

Asymptotic Minimavity, Optimal Posterior Concentration and Asymptotic Bayes Optimality of Horseshoe-type Priors Under Sparsity

Prasenjit Ghosh and Arijit Chakrabarti

Applied Statistics Unit, Indian Statistical Institute, Kolkata, India.

e-mail: prasenjit_r@isical.ac.in, arc@isical.ac.in

Abstract: In this article, we consider a general class of global-local scale mixtures of normals that is rich enough to include a wide variety of one-group shrinkage priors, such as the three parameter beta normal mixtures, the generalized double Pareto priors, and in particular, the horseshoe prior, the Strawderman–Berger prior and the normal–exponential–gamma priors. Asymptotic minimavity of Bayes estimates under the l_2 norm and posterior concentration properties corresponding to this general class of priors are studied when the data is assumed to be generated according to a multivariate normal distribution with a fixed unknown mean vector that is assumed to be sparse. Assuming that the number of non-zero means is known, we show that Bayes estimators based on this general class of shrinkage priors attain the minimax risk in the l_2 norm (up to some multiplicative constants). In particular, for the “horseshoe-type” priors, such as the horseshoe, the Strawderman–Berger and the standard double Pareto priors, the corresponding Bayes estimates are shown to attain the minimax l_2 risk asymptotically up to the correct constant. Moreover, it is shown that posterior distributions based on the chosen class of priors contract around the true mean vector at the minimax l_2 rate and around the corresponding Bayes estimates at least as fast as the minimax rate for a wide range of choices of the global shrinkage parameter. We also provide a lower bound to the total posterior spread corresponding to an important sub-family of the “horseshoe-type” priors, which gives important insight regarding the choice of the global shrinkage parameter for this sub-family of priors. This part of this article is inspired by the work of [van der Pas et al. \(2014\)](#). We come up with novel unifying arguments and extend their results over the general class of one-group priors under study. In the process, we settle a conjecture made in [Bhattacharya et al. \(2012\)](#) regarding such optimal posterior contraction properties of the normal–exponential–gamma and the generalized double Pareto priors.

In the second half of this article, we focus on the problem of simultaneous testing for the means of independent normal observations. We consider a Bayesian decision theoretic framework, where the overall loss is assumed to be the number of misclassified hypotheses and the data is modeled through a two-groups normal mixture distribution. We adopt the asymptotic framework of [Bogdan et al. \(2011\)](#) who introduced the notion of asymptotic Bayes optimality under sparsity (ABOS) in multiple hypothesis testing. The multiple testing procedures under study are induced by the general class of shrinkage priors considered in this paper and were also considered in [Ghosh et al. \(2015\)](#). Using one key technique essential for proving the aforesaid asymptotic minimavity result, we obtain exact asymptotic expressions of the Bayes risks of such induced decisions. A remarkable consequence of this is that such induced testing procedures based on the “horseshoe-type” priors are asymptotically Bayes optimal under sparsity (ABOS). This, to the best of our knowledge, is the first result in the literature where in a sparse problem, a class of one-group prior distributions is shown to achieve asymptotically the optimal two-groups solution up to the correct constant.

Keywords and phrases: Posterior concentration, minimavity, ABOS, three parameter beta, horseshoe, normal-exponential-gamma, Strawderman–Berger, generalized double Pareto.

1. Introduction

With rapid advancements in modern technology and computing facilities, high throughput data have become common in real life problems across diverse scientific fields such as genomics, biology, medicine, cosmology, finance, economics and climate studies etc. As a result, inferential problems involving a large number of unknown parameters are coming to the fore. Problems where the number of unknown parameters grows as least as fast as the number of observations are typically called high-dimensional. In such problems, often times only a few of these parameters are of real importance. For example, in a high-dimensional regression problem, the proportion of non-zero regressors or regressors with non-zero coefficients is often small compared to the total number of candidate regressors. This is called the phenomenon of sparsity. A common Bayesian approach to model data of this kind is to use a two-component point mass mixture prior for the parameters under consideration. Such priors put a positive mass at zero to model the null entries and a continuous distribution to identify the non-null effects. These are also referred to as “spike and slab priors” or two-groups priors. This is a very natural way of modeling data of this kind from a Bayesian view point. See [Johnstone and Silverman \(2004\)](#) and [Efron \(2004\)](#) in this context.

Use of the two-groups prior, although very natural, often poses a very daunting task computationally, especially in high-dimensional problems. Sometimes it is also possible that most of the parameters are very close to zero, but not exactly equal to zero. So in such a case a continuous prior may be able to capture sparsity in a more flexible manner. Due to these reasons, significant efforts have gone into modeling sparse high-dimensional data in recent times through hierarchical one-group continuous priors, which are also called one-group shrinkage priors. Bayesian analysis for such one-group priors are computationally much more tractable compared to the two-groups priors using MCMC techniques. A great variety of such one-group continuous shrinkage priors have appeared in the literature over the years. Some notable early examples are the t -prior in [Tipping \(2001\)](#), the double-exponential prior in [Park and Casella \(2008\)](#) and [Hans \(2009\)](#), and the normal-exponential-gamma priors in [Griffin and Brown \(2005\)](#). [Carvalho et al. \(2009, 2010\)](#) introduced the horseshoe prior, which has very appealing properties. Subsequently, many other one-group priors have been proposed in the literature, e.g, in [Polson and Scott \(2011, 2012\)](#), [Armagan et al. \(2011\)](#), [Armagan et al. \(2012\)](#) and [Griffin and Brown \(2010, 2012, 2013\)](#). The class of “three parameter beta normal” mixture priors was introduced in [Armagan et al. \(2011\)](#), while the “generalized double Pareto” class of priors was introduced by [Armagan et al. \(2012\)](#). The three parameter beta normal mixture family of priors contains among others the horseshoe, the Strawderman-Berger and the normal-exponential-gamma priors. A different class of one-group shrinkage priors referred to as the Dirichlet-Laplace (DL) priors has been introduced in [Bhattacharya et al. \(2012, 2014\)](#). Almost all such one-group priors can be expressed as multivariate scale-mixtures of normals by employing two levels of parameters to express the prior variances for the parameters under consideration, referred to as the “local shrinkage parameters” and a “global shrinkage parameter”. While the global shrinkage parameter accounts for the overall sparsity in the data, the local shrinkage parameters are helpful in detecting the obvious signals. Many of these priors are capable of handling sparsity by assigning a significant chunk of probability around zero, while at the same time they have tails which are heavy enough to ensure a priori large probabilities for the occurrence of large signals. As a result, noise observations are shrunk towards zero, while large observations almost remain unshrunk by such priors. The latter property is referred to as the “tail robustness” property and the corresponding one-group priors are called “tail robust” priors.

In recent times, researchers have started to investigate various optimality properties of estimators and testing rules based on one-group shrinkage priors, where the unknown mean or coefficient vectors are modeled through such hierarchical one-group formulation. [Datta and Ghosh \(2013\)](#) showed a near oracle optimality property of multiple testing rules due to [Carvalho et al. \(2010\)](#) based on the horseshoe prior in the context of multiple testing. [Ghosh et al. \(2015\)](#) extended the work of [Datta and Ghosh \(2013\)](#) for a broad class of one-group tail robust shrinkage priors that includes the horseshoe in particular. [Bhattacharya et al. \(2012, 2014\)](#) showed that for the estimation of a sparse multivariate normal mean vector under the quadratic loss function, the posterior arising from their proposed Dirichlet–Laplace prior contracts around the true mean vector at the minimax optimal rate with respect to the l_2 -norm for an appropriate choice of the underlying Dirichlet concentration parameter. In a recent article, [van der Pas et al. \(2014\)](#) showed that for the recovery of a sparse normal mean vector, the horseshoe estimator asymptotically achieves the minimax risk with respect to the l_2 loss up to a multiplicative constant and the corresponding posterior distribution contracts around the true mean vector at this minimax optimal rate. In addition, they showed that the posterior distribution based on the horseshoe prior contracts around the horseshoe estimator at least as fast as the minimax rate. See [Bickel et al. \(2009\)](#) and [Castillo et al. \(2014\)](#) where such issues were investigated for the Bayesian Lasso prior.

It is worth mentioning in this context that studying the concentration properties of posterior distributions has several important aspects in Bayesian inference. *Firstly*, it is often of interest to know whether a posterior distribution puts adequate mass around the neighborhood of the true distribution. For the sparse normal means model such as those considered in [Bhattacharya et al. \(2014\)](#) and [van der Pas et al. \(2014\)](#), one typically needs to show that the resulting posterior distribution puts probability mass 1 on balls centered around the true mean vector and having square radius proportional to the minimax risk under the l_2 norm. *Secondly*, for realistic uncertainty quantification, it is necessary that a posterior distribution contracts around its center at the same rate at which it converges to the true parameter value. See [van der Pas et al. \(2014\)](#) in this context. As commented in [Castillo and van der Vaart \(2012\)](#) that although construction of prior distributions in Bayesian inference is not driven by the ultimate goal of producing posterior distributions which have optimal concentration properties in terms of the frequentist minimax risk, for theoretical investigations, the minimax rate can be taken as a benchmark.

A question which is therefore natural to ask and also posed in [van der Pas et al. \(2014\)](#) is what aspects of one-group shrinkage priors are essential towards attaining optimal posterior concentration properties. [van der Pas et al. \(2014\)](#) also raised an important question that whether a sharp pick at the origin, combined with a thick tail, is essential for achieving such optimal concentration properties. [Bhattacharya et al. \(2012\)](#) conjectured that heavy tailed prior distributions, such as the horseshoe, the normal–exponential–gamma and the generalized double Pareto priors, should possess optimal posterior contraction properties in terms of the minimax l_2 risk. This has already been established for the horseshoe prior by [van der Pas et al. \(2014\)](#), while the case for the other two families of prior distributions remain unanswered. In an insightful article, [Polson and Scott \(2011\)](#) argued that, in sparse high-dimensional problems, one should choose the prior distribution corresponding to the local shrinkage parameters to be appropriately heavy-tailed so that large signals can escape the “gravitational pull” of the global variance component and are almost left unshrunk which is essential for the recovery and identification of true signals.

This motivates us to consider in the first half of this article, the problem of estimating a sparse multivariate normal mean vector based on a very general class of tail robust one-group prior distributions. The aforesaid general class is rich enough to include a wide variety of one-group shrinkage priors, such as the three parameter beta normal mixtures, the generalized double Pareto priors, the inverse-gamma priors, the half-t priors and many more. We work under the same framework as in [van der Pas et al. \(2014\)](#) and assume that the number of non-zero entries of the true mean vector is known. Our goal here is to study asymptotic risk properties of Bayes estimates based on the general class of one-group priors under consideration and also to investigate the concentration properties of the corresponding posterior distributions in terms of the minimax optimal rate under the l_2 norm. For that we take the global shrinkage parameter as a tuning parameter that we are free to choose depending on the proportion of non-null means. It is shown that when the underlying true mean vector is sparse in the nearly-black sense, the mean square errors of Bayes estimates based on such tail robust priors, are within a constant factor of the corresponding minimax risk with respect to the l_2 -norm, asymptotically. In particular, it is shown that for the “horseshoe-type” priors (to be defined in Section 2), such as, the three parameter beta normal mixtures with parameters $a = 0.5$, $b > 0$ (e.g. the horseshoe and the Strawderman–Berger priors), the generalized double Pareto priors with shape parameter $\alpha = 1$ (e.g. the standard double Pareto prior) and the inverse-gamma priors with shape parameter $\alpha = 0.5$, the corresponding Bayes estimates are asymptotically minimax in the sense that they asymptotically attain the minimax l_2 risk up to the correct constant, provided the global shrinkage parameter is asymptotically of the order of the proportion of non-null means or up to some logarithmic factor of it. This, to the best of our knowledge, is the first result in the literature which gives asymptotic minimavity property up to the correct constant of Bayes estimates arising out of such one-group priors. Till now, such result is known only for the horseshoe prior, and that too, only up to a multiplicative constant. This provides a useful guideline while deciding over the choice of such one-group priors for the recovery of sparse signals.

It is further shown that for a wide range of values of the global shrinkage parameter depending on the proportion of non-null means, the posterior distributions arising out of our general class of one-group priors under study, contract around the true mean vector at the minimax optimal rate under the l_2 norm and around the corresponding Bayes estimates at least as fast as this minimax rate for a broad range of choices of the global shrinkage parameter depending on the proportion of non-null means. Moreover, we also derive a lower bound to the total posterior spread corresponding to an important sub-family of the “horseshoe-type” priors, which provides important insight regarding the choice of the global shrinkage parameter for this sub-family of priors. A major contribution of our theoretical investigation is to show that shrinkage priors which are appropriately heavy-tailed, are able to attain the minimax optimal rate of contraction and one does not need a sharp peak at the origin, provided the global tuning parameter is chosen carefully. We provide some novel unifying arguments that work for this general class under study and extend the work of [van der Pas et al. \(2014\)](#) over this class. As an immediate consequence of our general theoretical results, we settle the conjecture of [Bhattacharya et al. \(2012\)](#) regarding such optimal posterior concentration properties of the normal–exponential–gamma and the generalized double Pareto priors.

One key technique used for proving the aforesaid asymptotic minimavity property turns out to be very handy in the study of asymptotic optimality (in a Bayesian decision theoretic frame-

work) of simultaneous hypothesis testing for independent normal means using one-group shrinkage priors. The works of [Datta and Ghosh \(2013\)](#) and [Ghosh et al. \(2015\)](#) already mentioned before established near asymptotic optimality properties of multiple testing rules based on one-group shrinkage priors when applied to data generated from a two-groups normal mixture model. Using an additive loss function, they showed that such multiple testing rules asymptotically attain the optimal Bayes risk for the corresponding two-groups formulation, up to a certain multiplicative factor. This indicates that the one-group solution can be a reasonable approximation asymptotically in the two-groups problem when the underlying mean vector is sparse. However, an interesting question that remains unanswered in these papers is whether such multiple testing procedures can asymptotically attain the optimal Bayes risk up to the correct constant. This question can be addressed satisfactorily by using the key technique mentioned before. We obtain exact asymptotic expressions of the Bayes risks of such multiple testing rules using some novel unifying arguments based on the aforesaid technique. It turns out that the answer is indeed in the affirmative for the “horseshoe-type” priors, whereby such decisions rules become asymptotically Bayes optimal under sparsity (ABOS), a notion that was introduced by [Bogdan et al. \(2011\)](#) in the context of multiple testing. This finding is remarkable since it is the first result of its kind where the one-group answer achieves exactly the optimal performance asymptotically when applied to a two-groups formulation. Moreover, it is theoretically established that for the present multiple testing problem, when the global shrinkage parameter is treated as a tuning parameter only, its optimal choice corresponding to the horseshoe-type priors, should be asymptotically of the order of the proportion of true alternatives, provided this proportion is known.

We organize the paper as follows. In Section 2, we describe the problem of estimating a sparse normal mean vector and the general class of tail robust shrinkage priors considered in this work. Section 3 contains the theoretical results involving asymptotic minimavity of Bayes estimators arising out of this general class of shrinkage priors under study and the corresponding posterior contraction results. Section 4 contains the theoretical results on asymptotic Bayes optimality under sparsity of the induced multiple testing procedures under study, followed by some concluding remarks in Section 5. Proofs of the main theorems and other supporting results are given in the Appendix.

Notations and Definition

We adopt the same convention of notation used in [van der Pas et al. \(2014\)](#) and [Ghosh et al. \(2015\)](#). Let $\{A_n\}$ and $\{B_n\}$ be any two sequences of non-negative real numbers indexed by n , such that, $B_n \neq 0$ for all sufficiently large n . We write $A_n \asymp B_n$ to denote $0 < \lim_{n \rightarrow \infty} \inf_n \frac{A_n}{B_n} \leq \lim_{n \rightarrow \infty} \sup_n \frac{A_n}{B_n} < \infty$ and $A_n \lesssim B_n$ to denote that there exists some constant $c > 0$ independent of n such that $A_n \leq cB_n$, provided n is sufficiently large. If $\lim_{n \rightarrow \infty} A_n/B_n = 1$ we write it as $A_n \sim B_n$. Moreover, if $|\frac{A_n}{B_n}| \leq D$ for all sufficiently n it is written as $A_n = O(B_n)$, where $D > 0$ is some positive real number independent of n , and if $\lim_{n \rightarrow \infty} A_n/B_n = 0$, it is expressed as $A_n = o(B_n)$. Thus $A_n = o(1)$ if $\lim_{n \rightarrow \infty} A_n = 0$. Given any two non-negative real valued functions $f(x)$ and $g(x)$, both having a common domain of definition (A, ∞) , $A \geq 0$, such that $g(x) \neq 0$ for all sufficiently large x , we write $f(x) \sim g(x)$ as $x \rightarrow \infty$ to denote $\lim_{x \rightarrow \infty} f(x)/g(x) = 1$.

Throughout this paper, Z is used to denote a $N(0, 1)$ random variable having cumulative distribution function and probability density function $\Phi(\cdot)$ and $\phi(\cdot)$, respectively.

Definition 1.1. A positive measurable function L defined over some (A, ∞) , $A \geq 0$, is said to be slowly varying (in Karamata's sense) if for each fixed $\alpha > 0$, $L(\alpha x) \sim L(x)$ as $x \rightarrow \infty$.

2. Estimation of a Sparse Normal Mean Vector and a General Class of One-Group Tail Robust Priors

Suppose that we observe an n -component random observation $(X_1, \dots, X_n) \in \mathbb{R}^n$, such that

$$X_i = \theta_i + \epsilon_i \text{ for } i = 1, \dots, n, \quad (2.1)$$

where the unknown parameters $\theta_1, \dots, \theta_n$ denote the effects under investigation and $\epsilon = (\epsilon_1, \dots, \epsilon_n) \sim N_n(0, I_n)$.

Let $l_0[q_n]$ denote the subset of \mathbb{R}^n given by,

$$l_0[q_n] = \{\theta \in \mathbb{R}^n : \#(1 \leq j \leq n : \theta_j \neq 0) \leq q_n\}. \quad (2.2)$$

Suppose we want to estimate the unknown mean vector $\theta = (\theta_1, \dots, \theta_n)$. For that we assume θ to be sparse in the “nearly black sense”, that is, $\theta \in l_0[q_n]$ with $q_n = o(n)$ as $n \rightarrow \infty$. Let $\theta_0 = (\theta_{01}, \dots, \theta_{0n})$ denote the true mean vector. Then the corresponding minimax rate with respect to the l_2 -norm is given by (see [Donoho et al. \(1992\)](#)),

$$\inf_{\hat{\theta}} \sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|\hat{\theta} - \theta_0\|^2 = 2q_n \log\left(\frac{n}{q_n}\right)(1 + o(1)), \text{ as } n \rightarrow \infty. \quad (2.3)$$

In (2.3) and throughout this paper E_{θ_0} denotes an expectation with respect to the $N_n(\theta_0, I_n)$ distribution. Our goal is to find an estimate of θ from a Bayesian view point having some good theoretical properties. As stated already in the introduction that a natural Bayesian approach to model (2.1) is to use a two-component point mass mixture prior for the θ_i 's, given by,

$$\theta_i \stackrel{i.i.d.}{\sim} (1 - p)\delta_{\{0\}} + p \cdot f, \quad i = 1, \dots, n. \quad (2.4)$$

where $\delta_{\{0\}}$ denotes the distribution having probability mass 1 at the point 0, and f denotes an absolutely continuous distribution over \mathbb{R} . Here the mixing proportion p is often interpreted as the theoretical proportion of non-null θ_i 's. See [Mitchell and Beauchamp \(1988\)](#) and [Johnstone and Silverman \(2004\)](#) in this context. It is usually recommended to choose a heavy tailed absolutely continuous distribution f over \mathbb{R} so that large observations can be recovered with higher degree of accuracy. [Johnstone and Silverman \(2004\)](#) used a t distribution in this context and used an empirical Bayes approach in order to estimate the unknown mixing proportion p via the method of marginal maximum likelihood and showed that if the coordinate-wise posterior median estimate is used, the resulting estimator attains the minimax rate with respect to the l_r loss, $r \in (0, 2]$. [Castillo and van der Vaart \(2012\)](#) studied the full Bayes approach where they found conditions on the two-groups prior that ensure contraction of the posterior distribution at the minimax rate. A comprehensive list of other empirical Bayes approaches using a two-groups model can be found in [Yuan and Lin \(2005\)](#), [Jiang and Zhang \(2009\)](#), [Castillo and van der Vaart \(2012\)](#), [Martin and Walker \(2014\)](#) and references therein.

As already mentioned in the introduction that although the two groups prior (2.4) is considered to be the most natural formulation for handling sparsity from a Bayesian view point, it offers a daunting computational challenge in high-dimensional problems. Due to this reason, the one-group formulation to model sparse data has received considerable attention from researchers over the years. Polson and Scott (2011) showed that almost all such shrinkage priors can be expressed as multivariate scale-mixture of normals. In this article, we consider Bayes estimators based on a general class of one-group shrinkage priors given through the following hierarchical one-group formulation:

$$\begin{aligned} X_i | \theta_i &\sim N(\theta_i, 1), \text{ independently for } i = 1, \dots, n \\ \theta_i | (\lambda_i^2, \tau^2) &\sim N(0, \lambda_i^2 \tau^2), \text{ independently for } i = 1, \dots, n \\ \lambda_i^2 &\sim \pi(\lambda_i^2), \text{ independently for } i = 1, \dots, n \end{aligned}$$

with $\pi(\lambda_i^2)$ being given by,

$$\pi(\lambda_i^2) = K(\lambda_i^2)^{-a-1} L(\lambda_i^2), \quad (2.5)$$

where $K \in (0, \infty)$ is the constant of proportionality, a is a positive real number and $L : (0, \infty) \rightarrow (0, \infty)$ is a measurable, non-constant slowly varying function. For the theoretical development in this paper, we assume that the function $L(\cdot)$ in (2.5) satisfies the following:

Assumption 2.1.

1. $\lim_{t \rightarrow \infty} L(t) \in (0, \infty)$, that is, there some exists $c_0(> 0)$ such that $L(t) \geq c_0$ for all $t \geq t_0$, for some $t_0 > 0$, which depends on both L and c_0 .
2. There exists some $0 < M < \infty$ such that $\sup_{t \in (0, \infty)} L(t) \leq M$.

Each λ_i^2 is referred to as a local shrinkage parameter and the parameter τ is called the global shrinkage parameter. For estimation of the unknown mean vector in the present normal means model (2.1), we treat τ as a tuning parameter that we are free to choose depending on the proportion of non-null means. From Theorem 1 of Polson and Scott (2011) it follows that the above general class of one-group priors will be “tail-robust” in the sense that for any given $\tau > 0$, $E(\theta_i | X_i, \tau) \approx X_i$, for large X_i ’s. We would like to mention here that a very broad class of one-group shrinkage priors actually fall inside this general class. Ghosh et al. (2015) established that the three parameter beta normal mixtures and the generalized double Pareto priors can be expressed in the above general form by showing that the corresponding prior distribution of the local shrinkage parameters can be written in the form given in (2.5) with the corresponding $L(\cdot)$ satisfying Assumption 2.1. It is easy to verify that some other well known shrinkage priors such as the families of inverse-gamma priors and the half-t priors are also covered by this general class of prior distributions under consideration. In case $a = 0.5$ for the general class of shrinkage priors under study, we refer the corresponding one-group priors as the *horseshoe-type* priors. The class of horseshoe-type priors encompasses an important sub-family of the general class of one-group priors under study, such as, the three parameter beta normal mixtures with parameters $a = 0.5$, $b > 0$ (e.g. the horseshoe, the Strawderman–Berger), the generalized double Pareto priors with shape parameter $\alpha = 1$ (e.g. the standard double Pareto), the inverse-gamma prior with shape parameter $\alpha = 0.5$ and many other shrinkage priors.

Now for a general global-local scale mixtures of normals, we have for each $i = 1, \dots, n$,

$$E(\theta_i | X_i, \tau) = (1 - E(\kappa_i | X_i, \tau)) X_i, \quad (2.6)$$

where $\kappa_i = 1/(1 + \lambda_i^2 \tau^2)$ is called the i -th posterior shrinkage coefficient. The resulting posterior mean $E(\theta|X, \tau) = (E(\theta_1|X_1, \tau), \dots, E(\theta_n|X_n, \tau))$ will be the Bayes estimate used to recover θ and will be denoted by $T_\tau(X)$. For notational convenience, we shall denote $E(\theta_i|X_i, \tau)$ by $T_\tau(X_i)$. It will be shown in the next section that when the true mean vector θ_0 is sparse in the “nearly black sense” and $a \in [0.5, 1)$, the estimator $T_\tau(X)$ will asymptotically attain the minimax rate (2.3) with respect to the l_2 norm, up to some multiplicative constant and that the posterior distribution contracts around the true θ_0 at this minimax optimal rate for suitably chosen values of τ depending on the proportion q_n/n . In particular, it will be shown that for the horseshoe-type priors, the resulting Bayes estimates will be asymptotically minimax in the sense that the ratio of the corresponding mean square error to the minimax risk in (2.3) asymptotically becomes 1 as the dimension n grows to infinity.

3. Asymptotic Minimavity and Posterior Contraction Rates

In this section, we present the theoretical results involving the mean square error for the Bayes estimates arising out of the general class of tail robust shrinkage priors under study, with $a \in [0.5, 1)$, and the spread of the corresponding posterior distributions. It is assumed that the number of non-zero components q_n of the unknown mean vector is known. Theorem 3.1 gives an upper bound on the mean square error for the Bayes estimates arising out of the general class one-group priors under consideration. Using this upper bound, it follows that for various choices of the global shrinkage parameter τ , depending on the proportion of non-zero means $\frac{q_n}{n}$, the aforesaid Bayes estimates attain the minimax risk with respect to the l_2 -norm up to some multiplicative constants. In particular, for the horseshoe-type priors, the corresponding Bayes estimates are shown to be asymptotically minimax. Theorem 3.2 gives an upper bound to the total posterior spread of our chosen one-group priors. Theorem 3.3 provides a sharp upper bound to the rate of contraction around the true mean vector for the posterior distributions arising out of the general class of shrinkage priors under study, which shows that such posterior distributions contract around the true mean vector at the minimax l_2 rate. Moreover, it also shows that these posterior distributions contract around the corresponding Bayes estimates at least as fast as the minimax optimal rate in the l_2 norm. Theorem 3.4 provides a lower bound to the total posterior variance for an important subclass of the horseshoe-type priors that gives more insight about the spread of such posterior distributions around the corresponding Bayes estimators for various choices of τ . Proofs of most of these results are given in the Appendix. Although we build on certain ideas of the proofs of the main theorems of van der Pas et al. (2014), we have to come up with novel unifying technical arguments that work for the kind of one-group priors studied in this paper. In particular, Lemmas A.3 - A.5, given in the Appendix, which form the crux of the arguments for proving the main theoretical results, namely, Theorems 3.1 - 3.3, are completely independent of the work of van der Pas et al. (2014). However, proof of Theorem 3.4 follows using some key arguments of van der Pas et al. (2014). Our work shows that some of the technical arguments used in van der Pas et al. (2014) can be used in greater generality.

Theorem 3.1. *Suppose $X \sim \mathcal{N}_n(\theta_0, I_n)$, where $\theta_0 \in l_0[q_n]$. Consider the general class of shrinkage priors where the prior distribution of the local shrinkage parameters λ_i^2 's is given by (2.5) with $0.5 \leq a < 1$ and the corresponding slowly varying component $L(\cdot)$ satisfies Assumption 2.1. Then*

the corresponding Bayes estimate $T_\tau(X)$, based on this general class of shrinkage priors, satisfies

$$\sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|T_\tau(X) - \theta_0\|^2 \lesssim q_n \log\left(\frac{1}{\tau^{2a}}\right) + (n - q_n) \tau^{2a} \sqrt{\log\left(\frac{1}{\tau^{2a}}\right)} \quad (3.1)$$

if $\tau \rightarrow 0$, as $n \rightarrow \infty$, $q_n \rightarrow \infty$ and $q_n = o(n)$.

Proof. See Appendix. ■

Using Theorem 3.1, it immediately follows that Bayes estimators arising out of the general class of shrinkage priors under study with $a \in [0.5, 1)$, attain the minimax risk with respect to the l_2 -norm, possibly up to some multiplicative factors, for various choices of the global shrinkage parameter τ . To see this, let us first fix any constant $c > 1$ and choose any $\rho > c$ in the proof of Theorem 3.1 (see Appendix). Then the corresponding multiplicative factor before the upper bound in (3.1) can at most be $4a\rho^2$. Now taking $\tau = (q_n/n)^\alpha$, $\alpha \geq 1$, or $\tau = (q_n/n)\sqrt{\log(n/q_n)}$ in (3.1), it follows that the corresponding mean square error can at most be of the order of $2q_n \log\left(\frac{n}{q_n}\right)$ up to the multiplicative factor $2a\rho^2 \max\{1, \alpha\}$, while it is always bounded below by the minimax l_2 -risk which is of the order $2q_n \log\left(\frac{n}{q_n}\right)$. Clearly,

$$\sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|T_\tau(X) - \theta_0\|^2 \asymp q_n \log\left(\frac{n}{q_n}\right). \quad (3.2)$$

However, a more interesting and remarkable consequence of Theorem 3.1 is the following. Consider the family of horseshoe-type priors for which one has $a = 0.5$ and let us take $\tau = q_n/n$ or $\tau = (q_n/n)\sqrt{\log(n/q_n)}$ in (3.1). Then using the proof of Theorem 3.1, we have,

$$\sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|T_\tau(X) - \theta_0\|^2 \leq 2\rho^2 q_n \log\left(\frac{n}{q_n}\right) (1 + o(1)) \text{ as } n \rightarrow \infty, \quad (3.3)$$

where the $o(1)$ term depends on $\rho > c > 1$. Now, using (3.3), and the fact that the minimax l_2 risk in (2.3) is the greatest lower bound to the mean square error term in (3.3), we obtain

$$1 \leq \frac{\sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|T_\tau(X) - \theta_0\|^2}{\inf_{\hat{\theta}} \sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|\hat{\theta} - \theta_0\|^2} \leq \rho^2 (1 + o(1)) \text{ as } n \rightarrow \infty. \quad (3.4)$$

Hence taking limit inferior and limit superior in (3.4) as $n \rightarrow \infty$, we have,

$$1 \leq \liminf_{n \rightarrow \infty} \frac{\sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|T_\tau(X) - \theta_0\|^2}{\inf_{\hat{\theta}} \sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|\hat{\theta} - \theta_0\|^2} \leq \limsup_{n \rightarrow \infty} \frac{\sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|T_\tau(X) - \theta_0\|^2}{\inf_{\hat{\theta}} \sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|\hat{\theta} - \theta_0\|^2} \leq \rho^2. \quad (3.5)$$

Note that the Bayes estimator $T_\tau(X)$ does not depend on the choice of $\rho > c > 1$. Hence the ratio in (3.4) is independent of the choice of $\rho > c > 1$. Consequently, the limit inferior and limit superior terms in (3.5) are both independent of how $\rho > c > 1$ are chosen. But choices of $\rho > c > 1$ in (3.5) are arbitrary. Therefore, taking infimum over all possible choices of $\rho > c > 1$ in (3.5) it follows that

$$\sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|T_\tau(X) - \theta_0\|^2 \sim \inf_{\hat{\theta}} \sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \|\hat{\theta} - \theta_0\|^2. \quad (3.6)$$

Observe that the above arguments also go through even if τ is taken to be asymptotically of the order of $\frac{q_n}{n}$ or $\frac{q_n}{n}\sqrt{\log(n/q_n)}$. Thus, (3.6) clearly shows that for carefully chosen values of the global tuning parameter τ depending on the proportion of non-null means $\frac{q_n}{n}$, Bayes estimators based on the horseshoe-type priors, asymptotically attains the minimax risk with respect to the l_2 -norm, up to the correct constant, and are therefore asymptotically minimax. It may be noted that it is a refinement of the corresponding result on asymptotic minimavity (up to a multiplicative constant) of the horseshoe estimator obtained by van der Pas et al. (2014). One possible explanation for such good performance of these priors is their ability to squelch the noise observations back to the origin, while leaving the large observations almost unshrunk, provided the global shrinkage parameter τ is carefully chosen. It is well known that smaller values of a typically result in a prior distribution with heavier tails. For example, the inverted beta families with $a = 0.5$ typically yields Cauchy like tails. See Polson and Scott (2012) in this context. Similar discussion on how the choice of the parameter a controls the tail behavior of the generalized double Pareto priors can also be found in Armagan et al. (2012). Recall that for the generalized double Pareto priors, we have $a = \alpha/2$, where α denotes the corresponding shape parameter (see Ghosh et al. (2015)). Armagan et al. (2012) argued for using smaller values of α to ensure strong shrinkage of noise-like observations towards the origin and to avoid the occurrence of large bias terms due to large signals. They recommended the standard double Pareto distribution as a default prior specification which has Cauchy like tail and for which one has $a = 0.5$. It is also worth noting in this context that Armagan et al. (2011) recommended using $a \in (0, 1)$ and $b \in (0, 1)$ for their proposed three parameter beta normal mixture priors. Their range of b is fully covered, while that of a is partially covered by Theorem 3.1. Thus, Theorem 3.1 provides strong theoretical support in favor of the recommendations of Armagan et al. (2011) and Armagan et al. (2012) regarding the choice of the hyperparameters in the corresponding one-group formulation.

Remark 3.1. Note that the aforesaid asymptotic minimavity property of Bayes estimates based on our chosen class of priors depends on treating τ as a tuning parameter to be chosen carefully depending on the proportion of non-zero means. Armagan et al. (2011) argued that the choice of the global tuning parameter should reflect the prior knowledge about the underlying sparsity presented in the data, provided such information is available. However, in practice one often doesn't have such information. In such situations, van der Pas et al. (2014) provided conditions under which the horseshoe estimator combined with an empirical Bayes estimate of the global variance component still attains the minimax risk when $q_n \propto n^\beta$, for $0 < \beta < 1$. We record here that we have a partly independent argument for proving such asymptotic minimavity result based on the empirical Bayes estimate proposed by van der Pas et al. (2014). This argument works for the general class of priors under study in the current paper. However, considering the length of the article, we omit the proof.

The next theorem gives an upper bound to the total posterior variance corresponding to our general class of heavy tailed shrinkage priors when $a \in [0.5, 1)$.

Theorem 3.2. Suppose $X \sim \mathcal{N}_n(\theta_0, I_n)$, where $\theta_0 \in l_0[q_n]$. Consider the general class of shrinkage priors where the prior distribution of the local shrinkage parameters λ_i^2 's is given by (2.5) with $0.5 \leq a < 1$ and the corresponding slowly varying component $L(\cdot)$ satisfies Assumption 2.1. Then the total posterior variance corresponding to this general class of shrinkage priors satisfies

$$\sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \sum_{i=1}^n \text{Var}(\theta_{0i}|X_i) \lesssim q_n \log\left(\frac{1}{\tau^{2a}}\right) + (n - q_n)\tau^{2a} \sqrt{\log\left(\frac{1}{\tau^{2a}}\right)}. \quad (3.7)$$

if $\tau \rightarrow 0$, as $n \rightarrow \infty$, $q_n \rightarrow \infty$ and $q_n = o(n)$.

Proof. See Appendix. ■

The next theorem gives upper bounds on the rate of contraction of posterior distributions based on our general class of one-group priors under study with $0.5 \leq a < 1$, both around the true mean vector as well as the corresponding Bayes estimates.

Theorem 3.3. *Under the assumptions of Theorem 3.2, if $\tau = \left(\frac{q_n}{n}\right)^\alpha$, $\alpha \geq 1$, or $\tau = \frac{q_n}{n} \sqrt{\log(n/q_n)}$, then*

$$\sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \Pi \left(\theta : \|\theta - \theta_0\|^2 > M_n q_n \log \left(\frac{n}{q_n} \right) | X \right) \rightarrow 0, \quad (3.8)$$

and

$$\sup_{\theta_0 \in l_0[q_n]} E_{\theta_0} \Pi \left(\theta : \|\theta - T_\tau(X)\|^2 > M_n q_n \log \left(\frac{n}{q_n} \right) | X \right) \rightarrow 0, \quad (3.9)$$

for every $M_n \rightarrow \infty$ as $n \rightarrow \infty$.

Proof. A straightforward application of Markov's inequality coupled with the results of Theorem 3.1 and Theorem 3.2 leads to (3.8), while (3.9) follows from the result of Theorem 3.2 together with the Markov's inequality. ■

Using (3.8) and (3.9), it follows that the posterior distributions based on our chosen class of one-group priors, with $0.5 \leq a < 1$, contract around both the true mean vector and the corresponding Bayes estimates at least as fast as the minimax l_2 risk in (2.3). On the other hand, Ghosal et al. (2000) showed that the posterior distributions cannot contract faster than the minimax risk around the truth. Hence, the rate of contraction of these posterior distributions around the true θ_0 must be the minimax optimal rate in (2.3), up to some multiplicative constants. However, the same may not be true for contraction around the corresponding Bayes estimates since no such result is available in the literature on the rates of contraction of posterior distributions around their respective posterior means. As already mentioned in the introduction that in order to provide a realistic measure of uncertainty, a posterior distribution needs to contract around the corresponding posterior mean at the same rate at which it approaches towards the true parameter value. Hence one needs to show that the posterior distributions in Theorem 3.3 contract around their corresponding Bayes estimates at the minimax optimal rate which cannot be established through (3.9). However, it seems that this requires a deeper investigation by invoking newer techniques and arguments to come up with a satisfactory answer to this question. We leave this as an important and interesting problem for future research.

In order to obtain a better insight regarding the spread of the posterior distribution around the Bayes estimates and the effect of choosing different values of τ on them, let us confine our attention to an important sub-family of the class of horseshoe-type prior distributions when the corresponding $L(\cdot)$ in (2.5) satisfies Assumption 2.1 and is non-decreasing over $(0, \infty)$. This sub-family of priors covers the three parameter beta normal mixtures with hyperparameters $a = 0.5$, $b > 0$ (e.g. the horseshoe, the Strawderman–Berger), the generalized double Pareto priors with shape parameter $\alpha = 1$ (e.g. the standard double Pareto), the inverse-gamma priors with shape parameter $\alpha = 0.5$ and many other such shrinkage priors. The next theorem gives a lower bound to the total posterior variance corresponding to this sub-family of priors which provides some useful insights into the effect of choosing τ depending on the proportion of non-zero means $\frac{q_n}{n}$.

Theorem 3.4. Suppose $X \sim \mathcal{N}_n(\theta_0, I_n)$, where $\theta_0 \in l_0[q_n]$. Consider the general class of shrinkage priors where the prior distribution of the local shrinkage parameters λ_i^2 's is given by (2.5) with $a = 0.5$, and the corresponding slowly varying component $L(\cdot)$ satisfies Assumption 2.1 and is non-decreasing over $(0, \infty)$. Then, the total posterior variance corresponding to the general class of shrinkage priors, satisfies

$$\sum_{i=1}^n E_{\theta_0} \text{Var}(\theta_{0i} | X_i) \gtrsim (n - q_n) \tau \sqrt{\log\left(\frac{1}{\tau}\right)}, \quad (3.10)$$

if $\tau \rightarrow 0$, as $n \rightarrow \infty$, $q_n \rightarrow \infty$ and $q_n = o(n)$.

Proof. See Appendix. ■

Like van der Pas et al. (2014), here also it can be observed from Theorem 3.4 that if we take $\tau = (q_n/n)^\alpha$ for $0 < \alpha < 1$, then the lower bound in (3.10) may exceed the minimax rate which implies that for such values of τ the corresponding prior distributions may have a sub-optimal posterior spread. Again, taking $\tau = q_n/n$ in Theorem 3.4, the corresponding lower bound in (3.10) is of the order $(q_n/n)\sqrt{\log(q_n/n)}$, thereby missing the minimax rate by the logarithmic factor $\sqrt{\log(q_n/n)}$. This also suggests that for $\tau = q_n/n$ such posterior distributions may contract around their respective centers at a rate faster than the minimax rate. Again, taking $\tau = (q_n/n)^\alpha$ with $\alpha \geq 1$ in the lower bound in Theorem 3.4 results a rate that is faster than the minimax l_2 risk, which means that for such choices of τ the corresponding lower bound in (3.10) fail to provide adequate information. However, if instead we choose $\tau = (\frac{q_n}{n})\sqrt{\log(n/q_n)}$ in Theorem 3.4, then using Theorem 3.2 and Theorem 3.4 it follows that the total posterior spread for this sub-family of priors is asymptotically of the order of the minimax l_2 risk in (2.3). Again, it has already been established that all the desired upper bounds in Theorem 3.1 and Theorem 3.2 are asymptotically of the order of the minimax error rate under the l_2 norm. Thus, like van der Pas et al. (2014), our theoretical results also suggest that for estimation of a sparse multivariate normal mean vector based on the horseshoe-type priors such as those in Theorem 3.4 and also for optimal posterior contraction properties, $\tau = (\frac{q_n}{n})\sqrt{\log(n/q_n)}$ should be regarded as the optimal choice of τ .

4. Asymptotic Bayes Optimality Under Sparsity

In the second part of this paper, we focus on the problem of simultaneous testing for the means of independent normal observations. For that we consider the normal means model in (2.1), where we have n independent observations X_1, \dots, X_n such that for each i , X_i is distributed according to a $N(\theta_i, 1)$ distribution. Suppose we wish to know whether for each i , θ_i is zero or not, that is, we wish to test $H_{0i} : \theta_i = 0$ against $H_{Ai} : \theta_i \neq 0$, for $i = 1, \dots, n$. Our focus is on situations when the unknown mean vector $(\theta_1, \dots, \theta_n)$ is sparse, that is, most of the null hypotheses H_{0i} 's are assumed to be true compared to the total number of tests n , which is typically assumed to be large. As already mentioned in the introduction that a natural Bayesian approach to formulate problems of this kind is to model the data through a two-component point mass mixture prior for the unknown θ_i 's. Let us introduce a set of latent indicator random variables ν_1, \dots, ν_n , where $\nu_i = 0$ denotes the event that H_{0i} is true while $\nu_i = 1$ corresponds to the event H_{0i} is false. Note that under H_{0i} , $\theta_i = 0$, while under H_{Ai} $\theta_i \neq 0$. So, let us assume that given $\nu_i = 0$, $\theta_i \sim \delta_{\{0\}}$, the distribution having probability mass 1 at the point 0, while $\theta_i | \nu_i = 1 \sim N(0, \psi^2)$, where $\psi^2 > 0$ is usually assumed to be large to accommodate the large signals. It is further assumed that the

unobservable indicator random variables ν_i 's are random samples from a Bernoulli(p) distribution, for some $p \equiv p_n$ in $(0, 1)$. The parameter p is often interpreted as the theoretical proportion of true alternatives. Marginalizing over the ν_i 's, given (p, ψ^2) , θ_i 's are assumed to be generated according to the following two-component point mass mixture distribution given by,

$$\theta_i \stackrel{i.i.d.}{\sim} (1-p)\delta_{\{0\}} + pN(0, \psi^2), \quad i = 1, \dots, n. \quad (4.1)$$

The marginal distribution of X_i 's is given by the following *two-groups* normal mixture model:

$$X_i \stackrel{i.i.d.}{\sim} (1-p)N(0, 1) + pN(0, 1 + \psi^2), \quad i = 1, \dots, n. \quad (4.2)$$

Under the above set up, the given testing problem now boils down to testing simultaneously

$$H_{0i} : \nu_i = 0 \text{ versus } H_{Ai} : \nu_i = 1 \text{ for } i = 1, \dots, n. \quad (4.3)$$

For the above multiple testing problem in (4.3), we consider a Bayesian decision theoretic framework described as follows. Let us assume that for each individual testing problem, the loss for committing a type I error and a type II error are the same and equal to 1 and the total loss is assumed to be the sum of losses incurred in each individual test. Thus the overall loss for the present multiple testing problem is the total number of misclassified hypotheses. Suppose t_{1i} and t_{2i} denote the probabilities of committing a type I error and a type II error respectively of a given multiple testing procedure for the i -th testing problem. Then, the corresponding Bayes risk of that procedure under the two-groups model (4.2), denoted R , is given by

$$R = \sum_{i=1}^n [(1-p)t_{1i} + pt_{2i}]. \quad (4.4)$$

Under this set up, [Bogdan et al. \(2011\)](#) showed that the Bayes rule which minimizes the Bayes risk in (4.4) is the test which, for each $i = 1, \dots, n$, declares the i -th null hypothesis H_{0i} to be significant if

$$\pi(\nu_i = 1 | X_i) > 0.5, \text{ or equivalently, } X_i^2 > c^2, \quad (4.5)$$

where

$$c^2 \equiv c_{\psi, f, \delta}^2 = \frac{1 + \psi^2}{\psi^2} (\log(1 + \psi^2) + 2 \log(\frac{1-p}{p})).$$

The above rule is also referred to as the Bayes Oracle by [Bogdan et al. \(2008\)](#) and [Bogdan et al. \(2011\)](#), since it involves the unknown mixing proportion p and the variance ψ^2 of the non-null θ_i 's which is also unknown, and hence cannot be attained for finite n .

By introducing two new parameters $u \equiv u_n = \psi_n^2$ and $v \equiv v_n = \psi_n^2 (\frac{1-p_n}{p_n})^2$, [Bogdan et al. \(2011\)](#) considered the following asymptotic scheme given by,

Assumption 4.1.

1. $p_n \rightarrow 0$ as $n \rightarrow \infty$.
2. $u_n = \psi_n^2 \rightarrow \infty$ as $n \rightarrow \infty$.
3. $\frac{\log v_n}{u_n} \rightarrow C \in (0, \infty)$ as $n \rightarrow \infty$.

Under Assumption 4.1, Bogdan et al. (2011) obtained the following asymptotic expressions for the probabilities of type I and type II errors corresponding to the Bayes Oracle (4.5), given by,

$$t_1^{BO} = e^{-C/2} \sqrt{\frac{2}{\pi v \log v}} (1 + o(1)), \text{ and} \quad (4.6)$$

$$t_2^{BO} = (2\Phi(\sqrt{C}) - 1)(1 + o(1)), \quad (4.7)$$

and the corresponding optimal Bayes risk, denoted R_{Opt}^{BO} , is given by,

$$R_{Opt}^{BO} = n((1-p)t_1^{BO} + pt_2^{BO}) = np(2\Phi(\sqrt{C}) - 1)(1 + o(1)). \quad (4.8)$$

Under this set up, Bogdan et al. (2011) introduced the notion of asymptotic Bayes optimality under sparsity (ABOS) defined as follows.

Definition 4.1. A multiple testing procedure with Bayes risk R with respect to the two-groups model (4.2), is said to be ABOS, if

$$\frac{R}{R_{Opt}^{BO}} \rightarrow 1 \text{ as } n \rightarrow \infty, \quad (4.9)$$

where R_{Opt}^{BO} denotes the optimal Bayes risk as given by (4.8) and the sequence of vectors $\{(\psi_n^2, p_n)\}_{n \geq 1}$ is assumed to satisfy the conditions of Assumption 4.1.

In other words, a multiple testing procedure is said to be ABOS if within the asymptotic framework Assumption 4.1, it attains the optimal Bayes risk in (4.8) up to the correct constant when the number of tests grows to infinity. Bogdan et al. (2011) provided necessary and sufficient conditions for a fixed threshold multiple testing rules to be ABOS and established that the step-up multiple testing procedure due to Benjamini and Hochberg (1995) and the Bonferroni procedure are ABOS under very general conditions.

For the independent normal means testing problem, Carvalho et al. (2010) observed through simulations that, under the assumption of sparsity, the posterior probability of the i -th alternative hypothesis being true under their two-groups formulation, can be well approximated by the posterior shrinkage weight $1 - \hat{\kappa}_i$, where $\hat{\kappa}_i$ denotes the i -th posterior shrinkage coefficient based on the horseshoe prior. Looking at the close proximity of these two posterior quantities, Carvalho et al. (2010) proposed a natural thresholding rule under a symmetric 0 – 1 loss based on the horseshoe prior, given by:

$$\text{reject } H_{0i} \text{ if } 1 - \hat{\kappa}_i > 0.5, \ i = 1, \dots, n.$$

Carvalho et al. (2010) numerically observed that under the assumption of sparsity, their proposed multiple testing procedure closely mimics the performance of the corresponding optimal Bayes rule based on their two-groups formulation which was later theoretically validated by Datta and Ghosh (2013). For the multiple testing problem as in (4.3), Datta and Ghosh (2013) considered the following decision rule based on the horseshoe prior assuming a symmetric 0 – 1 loss, given by

$$\text{reject } H_{0i} \text{ if } 1 - E(\kappa_i | X_i, \tau) > 0.5, \ i = 1, \dots, n, \quad (4.10)$$

where the data is assumed to be generated according to the two-groups model (4.2). They showed that within the asymptotic framework of Bogdan et al. (2011), if $\tau \sim p$, the induced decisions in

(4.10) attains the optimal Bayes risk in (4.8) up to a multiplicative constant, the constant being close to 1. In a more recent article, Ghosh et al. (2015) generalized and improved the results of Datta and Ghosh (2013) over a broad class of one-group shrinkage priors that includes the horseshoe in particular. They considered a general family of one-group tail robust prior distributions where the prior distribution $\pi(\lambda_i^2)$ for the local shrinkage parameters λ_i^2 is given by (2.5) and satisfies the following:

- (I) $\frac{1}{2} < a < 1$
- (II) $a = \frac{1}{2}$ and $L(t)/\sqrt{\log(t)} \rightarrow 0$ as $t \rightarrow \infty$.

Ghosh et al. (2015) showed that within their chosen asymptotic framework, if $\lim_{n \rightarrow \infty} \tau/p \in (0, \infty)$, then the Bayes risk of the multiple testing rules in (4.10) based on the above general class of prior distributions, denoted R_{OG} , satisfies

$$np[2\Phi(\sqrt{2a}\sqrt{C}) - 1](1 + o(1)) \leq R_{OG} \leq np[2\Phi\left(\sqrt{\frac{2aC}{\eta(1-\delta)}}\right) - 1](1 + o(1)) \text{ as } n \rightarrow \infty, \quad (4.11)$$

for every fixed $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$, where the $o(1)$ terms above are not necessarily the same, tend to zero as $n \rightarrow \infty$ and depend on the choice of $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$.

In case the proportion p is unknown, Ghosh et al. (2015) considered a data adaptive procedure by replacing τ by an empirical Bayes estimate $\hat{\tau}$ in the definition of the induced decisions in (4.10). The aforesaid empirical Bayes estimate of τ was proposed by van der Pas et al. (2014) and is given by,

$$\hat{\tau} = \max \left\{ \frac{1}{n}, \frac{1}{c_2 n} \sum_{j=1}^n 1\{|X_j| > \sqrt{c_1 \log n}\} \right\} \quad (4.12)$$

where $c_1 \geq 2$ and $c_2 \geq 1$ are some predetermined finite real numbers. Letting $E(1 - \kappa_i|X_i, \hat{\tau})$ denote the i -th posterior shrinkage weight $E(1 - \kappa_i|X_i, \tau)$ evaluated at $\tau = \hat{\tau}$, Ghosh et al. (2015) considered the following empirical Bayes procedure, given by,

$$\text{reject } H_{0i} \text{ if } 1 - E(\kappa_i|X_i, \hat{\tau}) > 0.5, \quad i = 1, \dots, n. \quad (4.13)$$

It was shown in Ghosh et al. (2015) that, within the asymptotic framework of Bogdan et al. (2011), if $p \equiv p_n \propto n^{-\beta}$, for $\beta \in (0, 1)$, the Bayes risk of the empirical Bayes multiple testing rules in (4.13) above, denoted R_{OG}^{EB} , is bounded above by,

$$R_{OG}^{EB} \leq np[2\Phi\left(\sqrt{\frac{2aC}{\eta(1-\delta)}}\right) - 1](1 + o(1)) \text{ as } n \rightarrow \infty, \quad (4.14)$$

for every fixed $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$, where the $o(1)$ term above tends to zero as $n \rightarrow \infty$ and depends on the choice of η and δ .

Thus, the induced decisions proposed by Carvalho et al. (2010) based on a general family of heavy tailed one-group prior distributions, asymptotically attain the Oracle risk (4.8) up to $O(1)$, with the constant in the one-group risk being close to that in the Oracle risk. However, an interesting question which naturally arises is whether it is possible for such induced decisions to attain

the optimal Bayes risk in (4.8) up to the correct constant, that is, whether such induced multiple testing rules can be asymptotically Bayes optimal when the hyperparameters (ψ_n^2, p_n) of the two-groups model satisfy Assumption 4.1. For that we consider in this section the multiple testing rules (4.10) and (4.13) imposed by our chosen general class of one-group shrinkage priors, where the prior distribution for the local shrinkage parameters λ_j^2 's is given by (2.5) and the corresponding slowly varying component $L(\cdot)$ satisfies Assumption 2.1. It will be seen in the forthcoming subsections that the answer to the aforesaid question of asymptotic Bayes optimality is indeed in the affirmative for the horseshoe-type prior distributions.

But before going into the theoretical details of this section, let us first have a look at where we gain in our present approach as compared to those in Datta and Ghosh (2013) and Ghosh et al. (2015). Note that, a careful inspection of the proofs of the last two articles shows that, for the induced decisions under study, the contribution due to erroneously rejecting the true nulls to the overall Bayes risk is negligible compared to that due to erroneous acceptance of false nulls. Hence, it is the contribution due to the type II errors only where one can improve upon. Note that, the authors of Datta and Ghosh (2013) and Ghosh et al. (2015) employed the following inequality:

$$E(\kappa_i|X_i, \tau) \leq \eta + \Pr(\kappa_i > \eta|X_i, \tau), \text{ for any } \eta \in (0, 1).$$

which played an important role for obtaining a non-trivial upper bound to the corresponding type II errors. On the other hand, in the present article, we use one key lemma involving the term $E(\kappa_i|X_i, \tau)$ (a version of which turns to be essential for proving the asymptotic minimavity and related contraction results of Section 3), followed by some novel and delicate arguments, which to the best of our knowledge, has not been reported elsewhere before. The aforesaid lemma describes an important asymptotic behavior of the posterior quantity $E(\kappa_i|x, \tau)$ for large x 's, when both x goes to infinity and τ tends to 0 simultaneously at an appropriate rate, which cannot be explained using the above inequality. This is where we differ distinctively in our present approach compared to those of Datta and Ghosh (2013) and Ghosh et al. (2015). Moreover, using the aforesaid lemma, we also obtain a sharp asymptotic lower bound to the corresponding type I error probabilities when τ is treated as a tuning parameter, which, in turn, helps to deduce the fact that for the horseshoe-type priors, the optimal choice of τ should be asymptotically of the order of p when it is treated as a tuning parameter only. This will be made more precise in Remark 4.1 of this paper.

4.1. Asymptotic Bounds on Error Probabilities of Both Kinds

Before studying the asymptotic risk properties of any multiple testing procedure, it is first necessary to investigate the asymptotic behaviors of the corresponding type I and type II error probabilities of the multiple testing procedure under study. In this section, we present four important results, namely, Theorem 4.1 - Theorem 4.4, involving asymptotic bounds to the probabilities of type I and type II errors of the individual induced decisions, both when τ is treated as a tuning parameter and it is replaced by the empirical Bayes estimate $\hat{\tau}$ defined in (4.12). While Theorem 4.1 and Theorem 4.2 below give asymptotic bounds to the error probabilities of both kinds of the i -th decision in (4.10), Theorem 4.3 and Theorem 4.4 give asymptotic upper bounds for the probabilities of type I and type II errors of the i -th empirical Bayes decision in (4.13), respectively. These results are of fundamental importance for analyzing the behavior of the multiple testing procedures (4.10) and

(4.13) in terms of their corresponding Bayes risks under the two groups normal mixture model (4.2) and the usual additive loss. It would be worth noting in this context that the asymptotic upper bound to the type I error probability (t_{1i}) as given in Theorem 4.1 and the asymptotic lower bound to the type II error probability (t_{2i}) as given by Theorem 4.1, are simple consequences of Theorem 4.4 and Theorem 4.7 of Ghosh et al. (2015), while the corresponding proofs for the asymptotic upper and lower bounds for t_{2i} and t_{1i} , respectively, require some novel arguments as already discussed at the end of the preceding section. Using this novel argument, we obtain sharp asymptotic upper and lower bounds to the probabilities of type II and type I errors of the i -th induced decision in (4.10) as given by Theorem 4.2 and Theorem 4.1, respectively, which cannot be improved further. This will be made more precise later in this paper. Moreover, proofs of Theorem 4.3 and Theorem 4.4 follow using Theorem 4.1 and Theorem 4.2, together with the arguments of Ghosh et al. (2015). Hence, we only state these results and excluded their proofs for the sake of brevity.

Theorem 4.1. *Suppose X_1, \dots, X_n are i.i.d. observations generated according to the two-groups normal mixture model (4.2) and suppose Assumption 4.1 is satisfied by the sequence of vectors (ψ^2, p) defining the two-groups model (4.2). Suppose we wish to test simultaneously $H_{0i} : \nu_i = 0$ vs $H_{Ai} : \nu_i = 1$, for $i = 1, \dots, n$, using the classification rule (4.10) induced by the general class of one-group shrinkage priors where the prior distribution of the local shrinkage parameter $\pi(\lambda_i^2)$ is given by (2.5) with $a \in (0, 1)$, and the corresponding slowly varying component $L(\cdot)$ satisfies Assumption 2.1. Suppose $\tau = \tau_n \rightarrow 0$ as $n \rightarrow \infty$. Let us fix any $\eta \in (0, 1)$ and any $\delta \in (0, 1)$. Then the probability t_{1i} of type I error of the i -th decision in (4.10) satisfies*

$$G_1(a, \eta, \delta) \frac{(\tau^{2a})^{\frac{\zeta}{2}}}{\sqrt{\log(\frac{1}{\tau^2})}} (1 + o(1)) \leq t_1 \equiv t_{1i} \leq H_1(a, \eta, \delta) \frac{\tau^{2a}}{\sqrt{\log(\frac{1}{\tau^2})}} (1 + o(1)) \text{ as } n \rightarrow \infty,$$

for any fixed $\zeta > \frac{2}{\eta(1-\delta)}$, where the $o(1)$ terms above do not depend on i and are not equal. Moreover, the $o(1)$ terms appearing on the left hand side of the above inequality depends on the choice of $\eta \in (0, 1)$ and $\delta \in (0, 1)$, while the $o(1)$ term on the right hand side of the above inequality is independent of the choices of $\eta \in (0, 1)$ and $\delta \in (0, 1)$. Here $G_1(a, \eta, \delta)$ and $H_1(a, \eta, \delta)$ are some finite positive constants each being independent of both i and m , but depend on $a \in (0, 1)$, $\eta \in (0, 1)$ and $\delta \in (0, 1)$.

Proof. See Appendix. ■

Theorem 4.2. *Consider the set-up of Theorem 4.1. Let us assume that $\tau = \tau_n \rightarrow 0$ as $n \rightarrow \infty$ such that $\lim_{n \rightarrow \infty} \frac{\tau}{p^\alpha} \in (0, \infty)$, for $\alpha > 0$. Then, for every fixed $\eta \in (0, 1)$ and every fixed $\delta \in (0, 1)$, the probability t_{2i} of type II error of the i -th individual decision in (4.10) satisfies*

$$[2\Phi(\sqrt{2a\alpha}\sqrt{C}) - 1](1 + o(1)) \leq t_2 \equiv t_{2i} \leq [2\Phi(\sqrt{\zeta a\alpha}\sqrt{C}) - 1](1 + o(1)) \text{ as } n \rightarrow \infty,$$

for any fixed $\zeta > \frac{2}{\eta(1-\delta)}$, where the $o(1)$ terms above do not depend on i and are not identical. Moreover, the $o(1)$ term on the right hand side of the above inequality depends on the choice of $\eta \in (0, 1)$ and $\delta \in (0, 1)$, while the $o(1)$ term on the left hand side of the above inequality is independent of the choices of $\eta \in (0, 1)$ and $\delta \in (0, 1)$.

Proof. See Appendix. ■

Theorem 4.3. Suppose X_1, \dots, X_n are i.i.d. observations generated according to the two-groups normal mixture model (4.2) and suppose Assumption 4.1 is satisfied by the sequence of vectors (ψ^2, p) defining the two-groups model (4.2), with $p \propto n^{-\beta}$, for some $0 < \beta < 1$. Suppose we wish to test simultaneously $H_{0i} : \nu_i = 0$ vs $H_{Ai} : \nu_i = 1$, for $i = 1, \dots, n$, using the classification rule (4.13) induced by the general class of one-group shrinkage priors where the prior distribution of the local shrinkage parameter $\pi(\lambda_i^2)$ is given by (2.5) with $a \in (0, 1)$, and the corresponding slowly varying component $L(\cdot)$ satisfies Assumption 2.1. Then, the probability \tilde{t}_{1i} of type I error of the i -th empirical Bayes decision in (4.13) satisfies

$$\tilde{t}_{1i} \leq \frac{B_1^* \alpha_n^{2a}}{\sqrt{\log(\frac{1}{\alpha_n^2})}} (1 + o(1)) + \frac{1/\sqrt{\pi}}{n^{c_1/2} \sqrt{\log n}} + e^{-2(2 \log 2 - 1) \beta_0 n p (1 + o(1))} \text{ as } n \rightarrow \infty,$$

where the $o(1)$ terms above are independent of i . Here B_1^* and β_0 are some finite positive constants, each being independent of both i and n , while $\alpha_n = \Pr(|X_1| > \sqrt{c_1 \log n})$ depends on n only.

Theorem 4.4. Let us consider the set-up of Theorem 4.3. Then, for every fixed $\eta \in (0, 1)$ and every fixed $\delta \in (0, 1)$, the probability \tilde{t}_{2i} of type II error of the i -th empirical Bayes decision in (4.13) satisfies

$$\tilde{t}_{2i} \leq [2\Phi(\sqrt{\zeta a \sqrt{C}}) - 1] (1 + o(1)) \text{ as } n \rightarrow \infty,$$

for any fixed $\zeta > \frac{2}{\eta(1-\delta)}$. Here the $o(1)$ term is independent of i , but depends on the choices of $\eta \in (0, 1)$ and $\delta \in (0, 1)$.

4.2. Asymptotic Bayes Optimality Under Sparsity of Induced Testing Rules Based on Horseshoe-type Priors

In this section, we present in Theorem 4.5 the exact asymptotic order of the ratio of Bayes risk of the induced multiple testing procedure (4.10) under study to that of the optimal Bayes risk (4.8) for a wide range of values of the global shrinkage parameter τ depending on the proportion of true alternatives p , which immediately shows that, within the asymptotic framework of Bogdan et al. (2011), if τ is asymptotically of the order of p , the multiple testing rules imposed by the horseshoe-type priors are ABOS. For the data adaptive empirical Bayes procedure, since it is already known that the corresponding Bayes risk is within a constant factor of the Oracle risk in (4.8) asymptotically (see Ghosh et al. (2015)), we focus on the more interesting situation when $a = 0.5$, that is, we consider the horseshoe-type priors only in this case. In Theorem 4.6, we establish that, within the asymptotic framework of Bogdan et al. (2011), if $p \propto n^{-\beta}$, $0 < \beta < 1$, the empirical Bayes decisions (4.13) based on the horseshoe-type priors will be ABOS. Proof of Theorem 4.5 is based on the asymptotic bounds for the corresponding type I and type II error probabilities as given by Theorem 4.1 and Theorem 4.2, followed by certain subtle analytic arguments, while proof of Theorem 4.6 is based on analogous arguments together with the techniques used for proving Theorem 3.2 of Ghosh et al. (2015) and hence, it is omitted.

Theorem 4.5. Let X_1, \dots, X_n , be i.i.d. observations having the two-groups normal mixture distribution (4.2) where the sequence of vectors (ψ^2, p) satisfy the conditions of Assumption 4.1. Suppose we wish to test the n hypotheses $H_{0i} : \nu_i = 0$ vs $H_{Ai} : \nu_i = 1$, for $i = 1, \dots, n$, simultaneously, using the classification rule (4.10) induced by the general class of one-group shrinkage priors where

the prior distribution of the local shrinkage parameter $\pi(\lambda_i^2)$ is given by (2.5) with $a \in [0.5, 1)$, and the corresponding slowly varying component $L(\cdot)$ satisfies Assumption 2.1. It is further assumed that $\tau \rightarrow 0$ as $n \rightarrow \infty$ such that $\lim_{n \rightarrow \infty} \tau/p^\alpha \in (0, \infty)$, for $\alpha \geq 1$. Then, the Bayes risk of the multiple testing rules in (4.10), denoted R_{OG} , satisfies

$$\lim_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} = \frac{2\Phi(\sqrt{2a\alpha}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1}. \quad (4.15)$$

In particular, for $a = 0.5$ and $\alpha = 1$ we have,

$$\lim_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} = 1.$$

Proof. See Appendix. ■

Some important consequences of Theorem 4.5 are the following. *First of all*, it gives an exact asymptotic expression of the Bayes risk R_{OG} of the induced decisions in (4.10), for any $a \in [0.5, 1)$ and any $\alpha \geq 1$. *Secondly*, it shows that for various choices of the global shrinkage parameter τ , the induced decisions based on our general class of one-group priors, asymptotically attain the optimal Bayes risk in (4.8) up to a multiplicative factor, the multiplicative factor being close to 1, provided α is not too large compared to 1. *Thirdly and most importantly*, it says that, if the global shrinkage parameter τ is asymptotically of the order of the proportion of true alternatives p , the induced decisions (4.10) based on the horseshoe-type priors (for which $a = 0.5$), asymptotically attain the optimal Bayes risk up to the correct constant, and hence, are ABOS. This not only sharpens the results of Datta and Ghosh (2013) and Ghosh et al. (2015), but at the same time provides an exact optimality result for a very broad class of one-group shrinkage priors in the context of multiple testing. Moreover, the above limiting expression in (4.15) clearly shows that the limiting value of the ratio of Bayes risks is an increasing function of both $\alpha \geq 1$ and $a \in [0.5, 1)$. Therefore, taking $\alpha = 1$ and $a = 0.5$ in (4.15) asymptotically yields the smallest possible value of the ratio of such Bayes risks, namely, 1. This explains why the hyperparameter a in the definition of $\pi(\lambda_i^2)$ in (2.5), should be set at $a = 0.5$ as a default choice compared to other values of $a \in (0.5, 1)$ for the present multiple testing problem.

Remark 4.1. Some important observations are to be made in this regard about the effect of choosing different values of τ depending on p , when p is assumed to be known. Although some of these observations were already made in Ghosh et al. (2015), the present remark helps us understand such facts very clearly. For that we confine our attention to the class of horseshoe-type priors for which one has $a = 0.5$. Note that, the type I and type II error probabilities of the i -th decision in (4.10), that is, t_{1i} and t_{2i} , do not depend on i and their common values are given by t_1 and t_2 , respectively. Thus, the Bayes risk of the decision rules in (4.10) is given by $R_{OG} = mp(\frac{1-p}{p}t_1 + t_2)$ (using (4.4)). Let us now assume that τ is asymptotically of the order of p^α , for some $\alpha > 0$. Let us first consider the case when $0 < \alpha < 1$. Then, given $\alpha \in (0, 1)$, there always exist some $\eta \in (0, 1)$ and some $\delta \in (0, 1)$, such that $0 < \alpha < \eta(1 - \delta)$. Let us now fix any $\zeta > 2/(\eta(1 - \delta))$, such that $\alpha < 2/\zeta < \eta(1 - \delta)$, that is, $\zeta \in (\frac{2}{\eta(1-\delta)}, \frac{2}{\alpha})$. Under this condition, using Theorem 4.1, we obtain the following:

$$\frac{t_1}{p} \gtrsim \frac{p^{\zeta\alpha/2-1}}{\sqrt{\log \frac{1}{p}}} \rightarrow \infty \text{ if } p_n \rightarrow 0 \text{ as } n \rightarrow \infty,$$

whence we have $R_{OG}/R_{Opt}^{BO} \rightarrow \infty$ as $n \rightarrow \infty$ within the asymptotic framework of Bogdan et al. (2011). Thus, the aforesaid asymptotic Bayes optimality result fails to hold in such situations. Consequently, for the present multiple testing problem, values of τ such that $\lim_{n \rightarrow \infty} \tau/p^\alpha \in (0, \infty)$, for $0 < \alpha < 1$, are not recommended for the horseshoe-type priors. On the other hand, the limiting value of the ratio R_{OG}/R_{Opt}^{BO} as given by (4.15), is non-decreasing in $\alpha \geq 1$, thereby attaining its minimum at $\alpha = 1$. Using the preceding observations, it follows that, for the horseshoe-type priors, the optimal choice of τ is asymptotically of the order of p , when the latter is assumed to be known. In all other cases where $0.5 < \alpha < 1$, though this choice of τ is not optimal in a strict mathematical sense, it still yields a Bayes risk that is within a constant factor of the Oracle risk (4.8) asymptotically, the factor being close to 1. Hence, for the present multiple testing problem, τ asymptotically of the order of p should be regarded as the optimal choice of τ when p is assumed to be known. It should however be noted in this context that the optimal choices of τ for the horseshoe-type priors such as those considered in Theorem 3.4 of Section 3, are not the same for simultaneous testing and estimation of independent normal means, and differ by a logarithmic factor of the proportion of non-zero means only.

The next theorem shows that, within the asymptotic framework of Bogdan et al. (2011), when $p_n \propto n^{-\beta}$ for $0 < \beta < 1$, the empirical Bayes induced testing procedures (4.13) based on the horseshoe-type priors asymptotically attain the optimal Bayes risk in (4.8) up to the correct constant, and hence, are ABOS. As already mentioned before, proof of this theorem follows using the arguments used in the proofs of Theorem 3.2 of Ghosh et al. (2015) and Theorem 4.5 of the present paper, and hence it is skipped.

Theorem 4.6. *Suppose X_1, \dots, X_n are i.i.d. observations generated according to the two-groups normal mixture model (4.2) and suppose Assumption 4.1 is satisfied by the sequence of vectors (ψ^2, p) defining the two-groups model (4.2), with $p \propto n^{-\beta}$, for some $0 < \beta < 1$. Suppose we wish to test simultaneously $H_{0i} : \nu_i = 0$ vs $H_{Ai} : \nu_i = 1$, for $i = 1, \dots, n$, using the classification rule (4.13) induced by the general class of one-group shrinkage priors where the prior distribution of the local shrinkage parameter $\pi(\lambda_i^2)$ is given by (2.5) with $a = 0.5$, and the corresponding slowly varying component $L(\cdot)$ satisfies Assumption 2.1. Then the Bayes risk of the empirical Bayes testing procedure (4.13), denoted R_{OG}^{EB} , satisfies*

$$\lim_{n \rightarrow \infty} \frac{R_{OG}^{EB}}{R_{Opt}^{BO}} = 1, \quad (4.16)$$

that is, the corresponding empirical Bayes decisions (4.13) will be ABOS.

5. Discussion

We studied in this paper various theoretical properties of a general class of heavy-tailed continuous shrinkage priors in terms of the quadratic minimax risk for estimating a multivariate normal mean vector which is known to be sparse in the sense of being nearly black. It is shown that Bayes estimators arising out of this general class asymptotically attain the minimax risk with respect to the l_2 norm, possibly up to some multiplicative constants. In particular, it is shown that for the horseshoe-type priors (defined in Section 2) such as the horseshoe, the Strawderman–Berger and the standard double Pareto priors, the corresponding Bayes estimators are asymptotically minimax in the sense that they attain the corresponding minimax risk under the l_2 -norm up to the correct

constant. Optimal rate of posterior contraction of these prior distributions around the truth in terms of the corresponding quadratic minimax error rate has also been established. We provided a novel unifying theoretical treatment that holds for a very broad class of one-group shrinkage priors. Another major contribution of this work is to show that shrinkage priors which are appropriately heavy tailed are good enough in order to attain the minimax optimal rate of contraction and that one doesn't need a pole at the origin, provided that the global tuning parameter is carefully chosen. This provides a partial answer to the question raised in [van der Pas et al. \(2014\)](#) already discussed in the introduction. As we already mentioned in Section 3 that one possible reason for such good performance of the kind of one-group shrinkage priors studied in this paper, is their ability to shrink the noise observations back to the origin, while leaving the large signals mostly unshrunk. Moreover, choice of the hyperparameter a also plays a significant role for optimal posterior contraction of these priors. We believe that the theoretical results in this paper can be extended further for a more general class of one-group priors through exploiting properties of general slowly varying functions such as those considered in [Ghosh et al. \(2015\)](#). It would be worth mentioning that the concentration and moment inequalities of [Ghosh et al. \(2015\)](#) and results like Lemma A.3 or Lemma A.7 of the present paper, are extremely useful for analyzing and understanding various theoretical properties of such one-group shrinkage priors, and these results form the basis of both the asymptotic minimavity and asymptotic Bayes optimality of the horseshoe-type priors under the assumption of sparsity. This makes us hopeful that the techniques employed in the present article would prove to be important ingredients for optimality studies of one-group continuous shrinkage priors.

In the latter half of this paper, we also studied the asymptotic risk properties of induced decisions based on our general class of continuous shrinkage priors in the context of multiple testing within a Bayesian decision theoretic framework. A major theoretical contribution of this work is to show that within the asymptotic scheme of [Bogdan et al. \(2011\)](#), such induced decisions based on the horseshoe-type priors become asymptotically Bayes optimal under sparsity. To the best of our knowledge, this is the first such result in the Bayesian literature where the two-groups answer can be exactly achieved asymptotically by an one-group formulation under the assumption of sparsity. Another important contribution of the present work is to theoretically establish the fact that, when $a = 0.5$, the optimal choice of the global variance component τ should be asymptotically of the order of the proportion of true alternatives p , when p is assumed to be known. Moreover, the present work also provides strong theoretical support in favor of using $a = 0.5$ as a default choice in our one-group prior specification. We hope that similar optimality results can be obtained for a more general class of one-group tail robust priors like those considered in [Ghosh et al. \(2015\)](#) for the present multiple testing problem by carefully exploiting properties of slowly varying functions, together with the powerful concentration and moment inequalities of [Ghosh et al. \(2015\)](#) and this way, one can relax the condition on the corresponding slowly varying component $L(\cdot)$ as given in Theorem 1 and Theorem 2 of [Ghosh et al. \(2015\)](#) for the case $a = 0.5$.

Over the past few years, one-group shrinkage priors have been gaining increasing popularity in the Bayesian literature for modeling sparse high-dimensional data instead of the more natural two-groups model. However, not much was known about their various theoretical properties until very recently. To the best of our knowledge, [Bhattacharya et al. \(2012\)](#) and [Datta and Ghosh \(2013\)](#) first studied certain asymptotic optimality properties of the Dirichlet–Laplace (DL) priors and the horseshoe prior, respectively, followed by the more recent works of [van der Pas et al. \(2014\)](#)

and Ghosh et al. (2015). We hope that the present work is a useful contribution towards that end and provides important theoretical justifications from both frequentist and Bayesian view point in favor of the use of such kind of one-group priors together with some useful guidelines regarding the choice of the underlying hyperparameters while deciding over a one-group formulation in a given problem. However, an interesting problem that remains open till date is to show asymptotic optimality properties of a full Bayes approach by assigning a hyperprior to the global shrinkage parameter τ . This applies in equal measure in both the problems of simultaneous testing and estimation. We hope to address this problem elsewhere in future.

Appendix

Lemma A.1. *For the general class of shrinkage priors (2.5) satisfying Assumption 2.1 the following holds true for any $0 < a < 1$:*

$$E(1 - \kappa | x, \tau) \leq \frac{KM}{a(1-a)} e^{\frac{x^2}{2}} \tau^{2a} (1 + o(1)), \text{ each fixed } x \in \mathbb{R},$$

where $\kappa = \frac{1}{1+\lambda^2\tau^2}$ denote the shrinkage coefficients and the $o(1)$ term depends only on τ such that $\lim_{\tau \rightarrow 0} o(1) = 0$.

Proof. See Ghosh et al. (2015). ■

Lemma A.2. *For every fixed $\tau > 0$, and each fixed $\eta, \delta \in (0, 1)$, the posterior distribution of the shrinkage coefficients $\kappa = 1/(1 + \lambda^2\tau^2)$ based on the general class of shrinkage priors (2.5) satisfying Assumption 2.1, with $a > 0$, satisfies the following concentration inequality:*

$$\Pr(\kappa > \eta | x, \tau) \leq \frac{H(a, \eta, \delta) e^{-\frac{\eta(1-\delta)x^2}{2}}}{\tau^{2a} \Delta(\tau^2, \eta, \delta)}, \text{ uniformly in } x \in \mathbb{R},$$

$$\text{where } \Delta(\tau^2, \eta, \delta) = \xi(\tau^2, \eta, \delta) L\left(\frac{1}{\tau^2} \left(\frac{1}{\eta\delta} - 1\right)\right),$$

$$\xi(\tau^2, \eta, \delta) = \frac{\int_{\frac{1}{\tau^2}(\frac{1}{\eta\delta}-1)}^{\infty} t^{-(a+\frac{1}{2}+1)} L(t) dt}{(a + \frac{1}{2})^{-1} \left(\frac{1}{\tau^2} \left(\frac{1}{\eta\delta} - 1\right)\right)^{-(a+\frac{1}{2})} L\left(\frac{1}{\tau^2} \left(\frac{1}{\eta\delta} - 1\right)\right)}, \text{ and}$$

$$H(a, \eta, \delta) = \frac{(a + \frac{1}{2})(1 - \eta\delta)^a}{K(\eta\delta)^{(a+\frac{1}{2})}},$$

where the term $\Delta(\tau^2, \eta, \delta)$ is such that $\lim_{\tau \rightarrow 0} \Delta(\tau^2, \eta, \delta)$ is a finite positive quantity for every fixed $\eta \in (0, 1)$ and every fixed $\delta \in (0, 1)$.

Proof. See Ghosh et al. (2015). ■

We now prove an extremely important lemma which together with one of its variants, namely, Lemma A.7, form the basis of all the major theoretical results deduced in this paper.

Lemma A.3. *Let us consider the general class of shrinkage priors (2.5) satisfying Assumption 2.1, with $a > 0$. Then, for each fixed $\tau \in (0, 1)$ and given any $c > 2$, the absolute difference between*

the Bayes estimators $T_\tau(x)$ based on the aforesaid class of shrinkage priors and an observation x , can be bounded above by a real valued function $h(\cdot, \tau)$, depending on c , and satisfying the following:

For any $\rho > c$,

$$\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho \log\left(\frac{1}{\tau^{2a}}\right)}} h(x, \tau) = 0.$$

Proof. By definition,

$$\begin{aligned} |T_\tau(x) - x| &= |xE(\kappa|x, \tau)| \\ &= \left| \frac{x \int_0^1 \kappa \cdot \kappa^{a+\frac{1}{2}-1} (1-\kappa)^{-a-1} L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa}-1\right)\right) e^{-\kappa x^2/2} d\kappa}{\int_0^1 \kappa^{a+\frac{1}{2}-1} (1-\kappa)^{-a-1} L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa}-1\right)\right) e^{-\kappa x^2/2} d\kappa} \right| \\ &= I(x, \tau), \text{ say.} \end{aligned}$$

Now observe that, for each fixed $\tau \in (0, 1)$, the function $|T_\tau(x) - x|$ is symmetric in x and it takes the value 0 when $x = 0$. Therefore, it would enough to find any non-negative function $h(x, \tau)$ that is symmetric in x and satisfies the stated conditions. Hence, without any loss of generality, let us assume that $x > 0$.

Let us fix any $\eta \in (0, 1)$ and any $\delta \in (0, 1)$.

Next, we observe that

$$I(x, \tau) \leq I_1(x, \tau) + I_2(x, \tau) \tag{A.1}$$

where $I_1(x, \tau) = |xE(\kappa 1\{\kappa < \eta\} | x, \tau)|$ and $I_2(x, \tau) = |xE(\kappa 1\{\kappa > \eta\} | x, \tau)|$.

Now using the variable transformation $t = \frac{1}{\tau^2}\left(\frac{1}{\kappa}-1\right)$, we have the following:

$$\begin{aligned} I_1(x, \tau) &= \left| \frac{x \int_0^\eta \kappa \cdot \kappa^{a+\frac{1}{2}-1} (1-\kappa)^{-a-1} L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa}-1\right)\right) e^{-\kappa x^2/2} d\kappa}{\int_0^1 \kappa^{a+\frac{1}{2}-1} (1-\kappa)^{-a-1} L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa}-1\right)\right) e^{-\kappa x^2/2} d\kappa} \right| \\ &= \left| \frac{x \int_{\frac{1}{\tau^2}(\frac{1}{\eta}-1)}^\infty \frac{1}{(1+t\tau^2)^{3/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt} \right| \\ &\leq \left| \frac{x \int_{\frac{1}{\tau^2}(\frac{1}{\eta}-1)}^\infty \frac{1}{(1+t\tau^2)^{3/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_{\frac{t_0}{\tau^2}}^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt} \right| \\ &= J_1(x, \tau) \text{ say,} \end{aligned} \tag{A.2}$$

Next observe that $\frac{t_0}{\tau^2} > t_0$ as $\tau^2 < 1$. Hence $L(t) \geq c_0$ for every $t \geq \frac{t_0}{\tau^2}$. Also, the function L is bounded by the constant $M > 0$. Utilizing these two observations and using the variable transformation $u = \frac{x^2}{1+t\tau^2}$ in both the numerator and the denominator of $J_1(x, \tau)$ in (A.2), and

writing $s = \frac{1}{1+t_0} \in (0, 1)$, we see that the term $J_1(x, \tau)$ can be bounded above as follows:

$$\begin{aligned} J_1(x, \tau) &\leq \frac{M}{c_0} \left| x \frac{\int_0^{\eta x^2} e^{-u/2} \left(\frac{u}{x^2}\right)^{3/2} \left(\frac{1}{\tau^2} \left(\frac{x^2}{u} - 1\right)\right)^{-a-1} \frac{x^2}{\tau^2 u^2} du}{\int_0^{sx^2} e^{-u/2} \left(\frac{u}{x^2}\right)^{1/2} \left(\frac{1}{\tau^2} \left(\frac{x^2}{u} - 1\right)\right)^{-a-1} \frac{x^2}{\tau^2 u^2} du} \right| \\ &= \frac{M}{c_0} \left| \frac{1}{x} \cdot \frac{\int_0^{\eta x^2} e^{-u/2} u^{a+3/2-1} \left(1 - \frac{u}{x^2}\right)^{-a-1} du}{\int_0^{sx^2} e^{-u/2} u^{a+1/2-1} \left(1 - \frac{u}{x^2}\right)^{-a-1} du} \right| \end{aligned}$$

Note that when $0 < u < \eta x^2$ we have $0 < \frac{u}{x^2} < \eta < 1$, that is, $1 - \eta < 1 - \frac{u}{x^2} < 1$. Similarly, we have $1 - s < 1 - \frac{u}{x^2} < 1$ when $0 < u < sx^2$. Using these observations we obtain,

$$\begin{aligned} J_1(x, \tau) &\leq \frac{M}{c_0(1-\eta)^{1+a}} \left| \frac{1}{x} \cdot \frac{\int_0^{\eta x^2} e^{-u/2} u^{a+3/2-1} du}{\int_0^{sx^2} e^{-u/2} u^{a+1/2-1} du} \right| \\ &\leq \frac{M}{c_0(1-\eta)^{1+a}} \left| \frac{1}{x} \cdot \frac{\int_0^\infty e^{-u/2} u^{a+3/2-1} du}{\int_0^{sx^2} e^{-u/2} u^{a+1/2-1} du} \right| \\ &= h_1(x, \tau) \text{ say,} \end{aligned} \tag{A.3}$$

where $h_1(x, \tau) = C_* \left[\left| x \int_0^{sx^2} e^{-u/2} u^{a+1/2-1} du \right| \right]^{-1}$ for some global constant $C_* \equiv C_*(a, \eta, L) > 0$ which is independent of both x and τ . Note that the function $h_1(x, \tau)$ is actually independent of τ and depends on x only.

Next we observe that,

$$\begin{aligned} I_2(x, \tau) &= |xE(\kappa 1\{\kappa > \eta\} \mid x, \tau)| \\ &\leq |x \Pr(\kappa > \eta \mid x, \tau)| \\ &\leq \left| x \frac{H(a, \eta, \delta) e^{-\frac{\eta(1-\delta)x^2}{2}}}{\tau^{2a} \Delta(\tau^2, \eta, \delta)} \right| \\ &= h_2(x, \tau) \text{ say,} \end{aligned} \tag{A.4}$$

Let $h(x, \tau) = h_1(x, \tau) + h_2(x, \tau)$. Therefore combining (A.1), (A.2), (A.3) and (A.4), we finally obtain for every $x \in \mathbb{R}$ and $\tau \in (0, 1)$,

$$|T_\tau(x) - x| \leq h(x, \tau). \tag{A.5}$$

Note that the function $h(x, \tau)$ defined above is symmetric in x about the origin. Now observe that the function $h_1(x, \tau)$ is strictly decreasing in $|x|$. Therefore, for any fixed $\tau \in (0, 1)$ and every $\rho > 0$,

$$\sup_{|x| > \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}} h_1(x, \tau) \leq C_* \left[\left| \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)} \int_0^{s\rho \log \left(\frac{1}{\tau^{2a}} \right)} e^{-u/2} u^{a+1/2-1} du \right| \right]^{-1}$$

implying that

$$\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}} h_1(x, \tau) = 0. \tag{A.6}$$

Again the function $h_2(x, \tau)$ is eventually decreasing in $|x|$. Therefore, for all sufficiently small $\tau \in (0, 1)$,

$$\sup_{|x| > \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}} h_2(x, \tau) \leq h_2\left(\sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}, \tau\right).$$

Let $\beta \equiv \beta(\eta, \delta) = \lim_{\tau \rightarrow 0} \Delta(\tau^2, \eta, \delta)$ for every fixed $\eta \in (0, 1)$ and every fixed $\delta \in (0, 1)$. Then $0 < \beta < \infty$ which follows from Lemma A.2. Then,

$$\begin{aligned} \lim_{\tau \rightarrow 0} h_2\left(\sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}, \tau\right) &= \frac{1}{\beta} \lim_{\tau \rightarrow 0} \left| \tau^{-2a} \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)} e^{-\frac{\eta(1-\delta)}{2} \rho \log \left(\frac{1}{\tau^{2a}} \right)} \right| \\ &= \frac{\sqrt{\rho}}{\alpha} \lim_{\tau \rightarrow 0} \left(\tau^{2a} \right)^{\frac{\eta(1-\delta)}{2} \left(\rho - \frac{2}{\eta(1-\delta)} \right)} \sqrt{\log \left(\frac{1}{\tau^{2a}} \right)} \\ &= \begin{cases} 0 & \text{if } \rho > \frac{2}{\eta(1-\delta)} \\ \infty & \text{otherwise,} \end{cases} \end{aligned}$$

whence it follows that

$$\lim_{\tau \rightarrow 0} \sup_{|x| > \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}} h_2(x, \tau) = \begin{cases} 0 & \text{if } \rho > \frac{2}{\eta(1-\delta)} \\ \infty & \text{otherwise.} \end{cases} \quad (\text{A.7})$$

Combining (A.6) and (A.7) together with the facts that

$$\lim_{\tau \rightarrow 0} \sup_{|x| > \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}} h(x, \tau) \leq \lim_{\tau \rightarrow 0} \sup_{|x| > \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}} h_1(x, \tau) + \lim_{\tau \rightarrow 0} \sup_{|x| > \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}} h_2(x, \tau)$$

and

$$\lim_{\tau \rightarrow 0} \sup_{|x| > \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}} h(x, \tau) \geq \lim_{\tau \rightarrow 0} \sup_{|x| > \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}} h_2(x, \tau)$$

it follows that

$$\lim_{\tau \rightarrow 0} \sup_{|x| > \sqrt{\rho \log \left(\frac{1}{\tau^{2a}} \right)}} h(x, \tau) = \begin{cases} 0 & \text{if } \rho > \frac{2}{\eta(1-\delta)} \\ \infty & \text{otherwise.} \end{cases} \quad (\text{A.8})$$

Observe that by choosing η appropriately close to 1 and δ close to 0, any real number larger than 2 can be expressed in the form $\frac{2}{\eta(1-\delta)}$. For example, taking $\eta = \frac{5}{6}$ and $\delta = \frac{1}{5}$ we obtain $\frac{2}{\eta(1-\delta)} = 3$. Hence, given any $c > 2$, let us choose $0 < \eta, \delta < 1$ such that $c = \frac{2}{\eta(1-\delta)}$. Clearly, the choice of $h(\cdot, \tau)$ depends on c . This, coupled with (A.5) and (A.8), completes the proof of Lemma A.3. \blacksquare

Remark A.1 The function $h(\cdot, \tau)$ defined in the proof of Lemma A.3 satisfies the following:

$$\lim_{|x| \rightarrow \infty} h(x, \tau) = 0, \text{ for each fixed } \tau \in (0, 1).$$

This means (using Lemma A.3) that for any fixed $\tau \in (0, 1)$, we have,

$$\lim_{|x| \rightarrow \infty} |T_\tau(x) - x| = 0. \quad (\text{A.9})$$

Equation (A.9) above shows that for the general class of tail robust priors under consideration, large observations almost remain unshrunk no matter however small τ is.

Proof of Theorem 3.1

Proof. Suppose that $X \sim \mathcal{N}_n(\theta, I_n)$, $\theta \in l_0[q_n]$ and $\tilde{q}_n = \#\{i : \theta_i \neq 0\}$. Let us split $E_\theta \|T_\tau(X) - \theta\|^2 = \sum_{i=1}^n E_{\theta_i} (T_\tau(X_i) - \theta_i)^2$ into two parts as follows:

$$\sum_{i=1}^n E_{\theta_i} (T_\tau(X_i) - \theta_i)^2 = \sum_{i: \theta_i \neq 0} E_{\theta_i} (T_\tau(X_i) - \theta_i)^2 + \sum_{i: \theta_i = 0} E_{\theta_i} (T_\tau(X_i) - \theta_i)^2 \quad (\text{A.10})$$

We shall show that, for all sufficiently small $\tau \in (0, 1)$, the mean square errors due to the non-zero means and the zero means as in the right hand side of (A.10), can be bounded above by $\tilde{q}_n \log\left(\frac{1}{\tau^{2a}}\right)$ and $(n - \tilde{q}_n)\tau^{2a} \sqrt{\log\left(\frac{1}{\tau^{2a}}\right)}$, respectively, up to some multiplicative constants.

Non-zero means:

For $\theta_i \neq 0$, let us split the corresponding mean square error as follows:

$$\begin{aligned} E_{\theta_i} (T_\tau(X_i) - \theta_i)^2 &= E_{\theta_i} ((T_\tau(X_i) - X_i) + (X_i - \theta_i))^2 \\ &= E_{\theta_i} (T_\tau(X_i) - X_i)^2 + E_{\theta_i} (X_i - \theta_i)^2 + 2E_{\theta_i} (T_\tau(X_i) - X_i)(X_i - \theta_i) \\ &\leq E_{\theta_i} (T_\tau(X_i) - X_i)^2 + 1 + 2\sqrt{E_{\theta_i} (T_\tau(X_i) - X_i)^2} \\ &= \left[\sqrt{E_{\theta_i} (T_\tau(X_i) - X_i)^2} + 1 \right]^2 \end{aligned} \quad (\text{A.11})$$

where we use the Cauchy-Schwartz inequality and the fact $E_{\theta_i} (X_i - \theta_i)^2 = 1$ in the step preceding the final step in the above chain of inequalities.

Let us now define

$$\zeta_\tau := \sqrt{2 \log\left(\frac{1}{\tau^{2a}}\right)}.$$

Let us now fix any $c > 1$ and choose any $\rho > c$. Then, using Lemma A.3, there exists a non-negative real-valued function $h(\cdot, \tau)$, depending on c , such that

$$|T_\tau(x) - x| \leq h(x, \tau), \text{ for all } x \in \mathbb{R} \quad (\text{A.12})$$

and

$$\lim_{\tau \downarrow 0} \sup_{|x| > \rho \zeta_\tau} h(x, \tau) = 0. \quad (\text{A.13})$$

Once again, using the fact $(T_\tau(X_i) - X_i)^2 \leq X_i^2$, together with (A.12), we obtain,

$$\begin{aligned} E_{\theta_i}(T_\tau(X_i) - X_i)^2 &= E_{\theta_i}[(T_\tau(X_i) - X_i)^2 1\{|X_i| \leq \rho\zeta_\tau\}] \\ &\quad + E_{\theta_i}[(T_\tau(X_i) - X_i)^2 1\{|X_i| > \rho\zeta_\tau\}] \\ &\leq \rho^2 \zeta_\tau^2 + \left(\sup_{|x| > \rho\zeta_\tau} h(x, \tau) \right)^2 \end{aligned} \quad (\text{A.14})$$

Now using (A.13) and the fact that $\zeta_\tau \rightarrow \infty$ as $\tau \rightarrow 0$, it follows that,

$$\left(\sup_{|x| > \rho\zeta_\tau} h(x, \tau) \right)^2 = o(\zeta_\tau^2) \text{ as } \tau \rightarrow 0. \quad (\text{A.15})$$

On combining (A.14) and (A.15), it follows that,

$$E_{\theta_i}(T_\tau(X_i) - X_i)^2 \leq \rho^2 \zeta_\tau^2 (1 + o(1)), \quad \text{as } \tau \rightarrow 0. \quad (\text{A.16})$$

Note that (A.16) holds uniformly for any i such that $\theta_i \neq 0$, whence we have,

$$\sum_{i: \theta_i \neq 0} E_{\theta_i}(T_\tau(X_i) - \theta_i)^2 \lesssim \tilde{q}_n \zeta_\tau^2, \quad \text{as } \tau \rightarrow 0. \quad (\text{A.17})$$

Zero means:

For $\theta_i = 0$, the corresponding mean square error is split as follows:

$$E_0[T_\tau(X_i)^2] = E_0[T_\tau(X_i)^2 1\{|X_i| \leq \zeta_\tau\}] + E_0[T_\tau(X_i)^2 1\{|X_i| > \zeta_\tau\}], \quad (\text{A.18})$$

where $\zeta_\tau = \sqrt{2 \log\left(\frac{1}{\tau^{2a}}\right)}$.

For the first term on the right hand side of (A.18), denoting $g_1(a) = \frac{KM}{a(1-a)}$ and using Lemma A.1, we obtain,

$$\begin{aligned} E_0 T_\tau(X_i)^2 1\{|X_i| \leq \zeta_\tau\} &\leq \frac{g_1(a)^2}{\sqrt{2\pi}} (\tau^{2a})^2 \int_{-\zeta_\tau}^{\zeta_\tau} x^2 e^{\frac{x^2}{2}} dx (1 + o(1)) \text{ as } \tau \rightarrow 0 \\ &= \frac{2g_1(a)^2}{\sqrt{2\pi}} (\tau^{2a})^2 \int_0^{\zeta_\tau} x^2 e^{\frac{x^2}{2}} dx (1 + o(1)) \text{ as } \tau \rightarrow 0 \\ &\leq \sqrt{\frac{2}{\pi}} g_1(a)^2 (\tau^{2a})^2 \zeta_\tau \frac{1}{\tau^{2a}} (1 + o(1)) \text{ as } \tau \rightarrow 0 \\ &\lesssim \zeta_\tau \tau^{2a} \text{ as } \tau \rightarrow 0, \end{aligned} \quad (\text{A.19})$$

while for the second term using the fact that $|T_\tau(x)| \leq |x|$ for $x \in \mathbb{R}$ we have,

$$\begin{aligned}
 E_0 T_\tau(X_i)^2 1\{|X_i| > \zeta_\tau\} &\leq 2 \int_{\zeta_\tau}^{\infty} x^2 \phi(x) dx \\
 &= 2[\zeta_\tau \phi(\zeta_\tau) + (1 - \Phi(\zeta_\tau))] \\
 &\leq 2\zeta_\tau \phi(\zeta_\tau) + 2 \frac{\phi(\zeta_\tau)}{\zeta_\tau} \\
 &= \sqrt{\frac{2}{\pi}} \zeta_\tau \tau^{2a} (1 + o(1)) \text{ as } \tau \rightarrow 0 \\
 &\lesssim \zeta_\tau \tau^{2a} \text{ as } \tau \rightarrow 0.
 \end{aligned} \tag{A.20}$$

Combining equations (A.18), (A.19) and (A.20), it follows that,

$$\sum_{i: \theta_i=0} E_{\theta_i} (T_\tau(X_i) - \theta_i)^2 \lesssim (n - \tilde{q}_n) \tau^{2a} \sqrt{\log\left(\frac{1}{\tau^{2a}}\right)} \text{ as } \tau \rightarrow 0. \tag{A.21}$$

Finally, on combining (A.10), (A.17) and (A.21), we have,

$$\sum_{i=1}^n E_{\theta_i} (T_\tau(X_i) - \theta_i)^2 \lesssim \tilde{q}_n \log\left(\frac{1}{\tau^{2a}}\right) + (n - \tilde{q}_n) \tau^{2a} \sqrt{\log\left(\frac{1}{\tau^{2a}}\right)} \text{ as } \tau \rightarrow 0. \tag{A.22}$$

The stated result now becomes immediate by observing that $\tilde{q}_n \leq q_n$ and $q_n = o(n)$ and then taking supremum over all $\theta \in l_0[q_n]$. This completes the proof of Theorem 3.1. \blacksquare

Lemma A.4. *The posterior variance arising out of the general class of shrinkage priors (2.5) can be represented by the following identity:*

$$\text{Var}(\theta|x) = \frac{T_\tau(x)}{x} - (T_\tau(x) - x)^2 + x^2 \frac{\int_0^\infty \frac{1}{(1+t\tau^2)^{5/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt} \tag{A.23}$$

$$= \frac{T_\tau(x)}{x} - T_\tau^2(x) + x^2 \frac{\int_0^\infty \frac{(t\tau^2)^2}{(1+t\tau^2)^{5/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt} \tag{A.24}$$

which can be bounded from above by

$$\text{Var}(\theta|x) \leq 1 + x^2.$$

Proof. By the law of iterated variance it follows that

$$\begin{aligned}
 \text{Var}(\theta|x) &= E[\text{Var}(\theta|x, \kappa, \tau)] + \text{Var}[E(\theta|x, \kappa, \tau)] \\
 &= E[(1 - \kappa)|x, \tau] + \text{Var}[x(1 - \kappa)|x, \tau] \\
 &= E[(1 - \kappa)|x, \tau] + x^2 \text{Var}[\kappa|x, \tau] \\
 &= E[(1 - \kappa)|x, \tau] + x^2 E[\kappa^2|x, \tau] - x^2 E^2[\kappa|x, \tau] \\
 &= \frac{T_\tau(x)}{x} - (T_\tau(x) - x)^2 + x^2 \frac{\int_0^\infty \frac{1}{(1+t\tau^2)^{5/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}
 \end{aligned}$$

which can equivalently be represented as by the following identity as well:

$$\begin{aligned} \text{Var}(\theta|x) &= E[(1 - \kappa)|x, \tau] + x^2 E[(1 - \kappa)^2|x, \tau] - x^2 E^2[1 - \kappa|x, \tau] \\ &= \frac{T_\tau(x)}{x} - T_\tau^2(x) + x^2 \frac{\int_0^\infty \frac{(t\tau^2)^2}{(1+t\tau^2)^{5/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}. \end{aligned}$$

That $\text{Var}(\theta|x) \leq 1 + x^2$ now follows trivially from the above identities. ■

Lemma A.5. *Let us define*

$$J(x, \tau) = x^2 \frac{\int_0^\infty \frac{(t\tau^2)^2}{(1+t\tau^2)^{5/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}, \quad x \in \mathbb{R} \text{ and } \tau \in (0, 1),$$

where the function $L(\cdot)$ is already defined in (2.5) and satisfies Assumption 2.1, with $a \in [0.5, 1)$. Then, for each $x \in \mathbb{R}$ and every $0 < \tau < 1$, the function $J(\cdot, \cdot)$ is bounded above by,

$$J(x, \tau) \leq 2KM e^{\frac{x^2}{2}} \tau^{2a} (1 + o(1)), \quad (\text{A.25})$$

where the $o(1)$ term is independent of x , and depends only on τ such that the term $(1 + o(1))$ in (A.25) is positive for any $0 < \tau < 1$, and $\lim_{\tau \rightarrow 0} o(1) = 0$.

Proof. First observe that, for each fixed $\tau \in (0, 1)$, the function $J(x, \tau)$ is symmetric in x , and so is the stated upper bound in (A.25). Moreover, $J(0, \tau) = 0$, for any $0 < \tau < 1$. Thus, the stated result is vacuously true when $x = 0$ and $0 < \tau < 1$. Therefore, it will be enough to prove that (A.25) holds when $x > 0$. So, let us assume that $x > 0$.

Note that

$$\begin{aligned} J(x, \tau) &\leq Mx^2 e^{\frac{x^2}{2}} \frac{\int_0^\infty \frac{(t\tau^2)^2}{(1+t\tau^2)^{5/2}} t^{-a-1} e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-a-1} L(t) dt} \\ &= KMx^2 e^{\frac{x^2}{2}} \int_0^\infty \frac{(t\tau^2)^2}{(1+t\tau^2)^{5/2}} t^{-a-1} e^{-\frac{x^2}{2(1+t\tau^2)}} dt (1 + o(1)) \end{aligned} \quad (\text{A.26})$$

where in the preceding chain of inequalities we use the facts that the function $L(\cdot)$ is bounded above by the constant $M > 0$ and $\int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-a-1} L(t) dt = \int_0^\infty t^{-a-1} L(t) dt (1 + o(1)) = K^{-1} (1 + o(1))$ as $\tau \rightarrow 0$ (which follows from Lebesgue's Dominated Convergence Theorem). Clearly, the $o(1)$ term does not involve x and depends only on τ such that $\lim_{\tau \rightarrow 0} o(1) = 0$. It is also evident that the term $(1 + o(1))$ in (A.25) is always positive for any $0 < \tau < 1$.

Consider now the following variable transformation in the integral in (A.26):

$$u = \frac{x^2}{1 + t\tau^2}.$$

Then we have,

$$\begin{aligned} J(x, \tau) &\leq KMx^2 e^{\frac{x^2}{2}} \int_0^{x^2} \left(1 - \frac{u}{x^2}\right)^2 \left(\frac{u}{x^2}\right)^{1/2} \left(\frac{1}{\tau^2} \frac{x^2}{u} \left(1 - \frac{u}{x^2}\right)\right)^{-a-1} e^{-u/2} \frac{x^2}{\tau^2 u^2} du (1 + o(1)) \\ &= KM(x^2)^{1/2-a} e^{\frac{x^2}{2}} \tau^{2a} \int_0^{x^2} \left(1 - \frac{u}{x^2}\right)^{1-a} u^{a-1/2} e^{-u/2} du (1 + o(1)) \end{aligned} \quad (\text{A.27})$$

Note that $0 < u < x^2$ implies $0 < 1 - \frac{u}{x^2} < 1$. Hence, $(1 - \frac{u}{x^2})^{1-a} < 1$ as $0 < a < 1$. Also, as $\frac{1}{2} \leq a < 1$, we have, $u^{a-1/2} \leq (x^2)^{a-1/2}$.

Therefore using (A.27), we obtain,

$$\begin{aligned} J(x, \tau) &\leq KM e^{\frac{x^2}{2}} \tau^{2a} \int_0^{x^2} e^{-u/2} du (1 + o(1)), \\ &= 2KM e^{\frac{x^2}{2}} \tau^{2a} (1 - e^{-x^2/2}) (1 + o(1)) \\ &\leq 2KM e^{\frac{x^2}{2}} \tau^{2a} (1 + o(1)), \end{aligned}$$

thereby completing the proof of Lemma A.5. ■

Proof of Theorem 3.2

Proof. Suppose that $X \sim \mathcal{N}_n(\theta, I_n)$, where $\theta \in l_0[q_n]$ and $\tilde{q}_n = \#\{i : \theta_i \neq 0\}$. Then $\tilde{q}_n \leq q_n$. Let $\zeta_\tau = \sqrt{2 \log \left(\frac{1}{\tau^{2a}} \right)}$.

Nonzero means:

By applying the same reasoning as in the proof of Lemma A.3 to the final term of $\text{Var}(\theta|x)$ in (A.23), there exists a non-negative real-valued measurable function $\tilde{h}(x, \tau)$ such that $\text{Var}(\theta|x) \leq \tilde{h}(x, \tau)$ for all $x \in \mathbb{R}$ and all $\tau \in (0, 1)$, where $\tilde{h}(x, \tau) \rightarrow 1$ as $x \rightarrow \infty$ for any fixed $\tau \in (0, 1)$. If $\tau \rightarrow 0$, the function $\tilde{h}(x, \tau)$ satisfies the following for any $c > 1$:

$$\lim_{\tau \downarrow 0} \sup_{|x| > \rho \zeta_\tau} \tilde{h}(x, \tau) = 1 \text{ for all } \rho > c.$$

So, let us fix any arbitrary $c > 1$ and choose any $\rho > c$. Then using the above arguments it follows that,

$$E_{\theta_i} [\text{Var}(\theta_i | X_i) 1\{|X_i| > \rho \zeta_\tau\}] \lesssim 1 \text{ as } \tau \rightarrow 0. \quad (\text{A.28})$$

Let us now consider the case $|x| \leq \rho \zeta_\tau$. Then using the fact that $\text{Var}(\theta|x) \leq 1 + x^2$ for any $x \in \mathbb{R}$ as obtained from Lemma A.4, we obtain,

$$E_{\theta_i} [\text{Var}(\theta_i | X_i) 1\{|X_i| \leq \rho \zeta_\tau\}] \lesssim \zeta_\tau^2 \text{ as } \tau \rightarrow 0. \quad (\text{A.29})$$

Combining (A.28) and (A.29) together, it follows that, for any i such that $\theta_i \neq 0$,

$$\begin{aligned} &E_{\theta_i} \text{Var}(\theta_i | X_i) \\ &= E_{\theta_i} [\text{Var}(\theta_i | X_i) 1\{|X_i| > \rho \zeta_\tau\}] + E_{\theta_i} [\text{Var}(\theta_i | X_i) 1\{|X_i| \leq \rho \zeta_\tau\}] \\ &\lesssim 1 + \zeta_\tau^2 \text{ as } \tau \rightarrow 0, \end{aligned}$$

which holds uniformly in i such that $\theta_i \neq 0$. Thus,

$$\sum_{i:\theta_i \neq 0} E_{\theta_i} \text{Var}(\theta_i | X_i) \lesssim \tilde{q}_n (1 + \zeta_\tau^2) \text{ as } \tau \rightarrow 0. \quad (\text{A.30})$$

Zero means:

When $|x| > \zeta_\tau$, using the bound $\text{Var}(\theta|x) \leq 1 + x^2$ in Lemma A.4, we obtain,

$$\begin{aligned} E_0 \text{Var}(\theta | X_i) 1\{|X_i| > \zeta_\tau\} &\leq 2 \int_{\zeta_\tau}^{\infty} (1 + x^2) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &\lesssim \frac{\tau^{2a}}{\zeta_\tau} + \zeta_\tau \tau^{2a} \text{ as } \tau \rightarrow 0. \end{aligned} \quad (\text{A.31})$$

Again when $|x| \leq \zeta_\tau$, we consider the upper bound $\text{Var}(\theta|x) \leq \frac{T_\tau(x)}{x} + J(x, \tau)$ as obtained from Lemma A.4, where the term $J(x, \tau)$ is already defined in Lemma A.5. Also note that $\frac{T_\tau(x)}{x} = E(1 - \kappa|x, \tau)$. Therefore, using the moment inequality given in Lemma A.1, it follows

$$\int_{-\zeta_\tau}^{\zeta_\tau} \frac{T_\tau(x)}{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \lesssim \tau^{2a} \zeta_\tau \text{ as } \tau \rightarrow 0. \quad (\text{A.32})$$

Similarly, using Lemma A.5 we obtain,

$$\int_{-\zeta_\tau}^{\zeta_\tau} J(x, \tau) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \lesssim \tau^{2a} \zeta_\tau \text{ as } \tau \rightarrow 0. \quad (\text{A.33})$$

On combining (A.32) and (A.33), it follows that

$$E_0 [\text{Var}(\theta_i | X_i) 1\{|X_i| \leq \zeta_\tau\}] \lesssim \tau^{2a} \zeta_\tau \text{ as } \tau \rightarrow 0. \quad (\text{A.34})$$

Therefore, using (A.31) and (A.34), it follows that for any i such that $\theta_i = 0$,

$$\begin{aligned} &E_0 \text{Var}(\theta_i | X_i) \\ &= E_0 [\text{Var}(\theta_i | X_i) 1\{|X_i| > \zeta_\tau\}] + E_0 [\text{Var}(\theta_i | X_i) 1\{|X_i| \leq \zeta_\tau\}] \\ &\lesssim \tau^{2a} \zeta_\tau \text{ as } \tau \rightarrow 0, \end{aligned}$$

which holds uniformly in i such that $\theta_i = 0$. Consequently,

$$\sum_{i:\theta_i=0} E_0 \text{Var}(\theta_i | X_i) \lesssim (n - \tilde{q}_n) \tau^{2a} \zeta_\tau \text{ as } \tau \rightarrow 0. \quad (\text{A.35})$$

Combining (A.30) and (A.35), it follows that

$$E_\theta \sum_{i=1}^n \text{Var}(\theta_i | X_i) \lesssim \tilde{q}_n \log\left(\frac{1}{\tau^{2a}}\right) + (n - \tilde{q}_n) \tau^{2a} \sqrt{\log\left(\frac{1}{\tau^{2a}}\right)} \text{ as } \tau \rightarrow 0.$$

The result then follows immediately by noting that $\tilde{q}_n \leq q_n$ and $q_n = o(n)$ and subsequently taking supremum over all $\theta \in l_0[q_n]$. This completes the proof of Theorem 3.2. \blacksquare

Lemma A.6. Suppose the function $L(\cdot)$ given by (2.5) satisfies Assumption 2.1 and is non-decreasing over $(0, \infty)$, with $a = \frac{1}{2}$. Let us define for fixed $y > 0$ and for fixed $k > 0$,

$$I_k = \int_0^\infty \frac{(t\tau^2)^{k-\frac{1}{2}}}{(1+t\tau^2)^k} t^{-3/2} L(t) e^{\frac{t\tau^2}{1+t\tau^2}y} dt.$$

Then,

$$\begin{aligned} I_{\frac{5}{2}} &\geq L(1)\tau \left[\frac{\tau}{y} (e^{y/2} - e^{\tau^2 y}) + \frac{1}{\sqrt{2}y} (e^y - e^{y/2}) \right], \text{ for } \tau < \frac{1}{\sqrt{2}} \\ I_{\frac{1}{2}} &\leq \tau \left[\frac{e^{\tau^2 y}}{K\tau} + 2Me^{\tau y} \left(\frac{1}{\tau} - \frac{1}{\sqrt{\tau}} \right) + 2Me^{\frac{y}{2}} \left(\frac{1}{\sqrt{\tau}} - \sqrt{2} \right) + \frac{2M\sqrt{2}}{y} (e^y - e^{\frac{y}{2}}) \right], \text{ for } \tau < \frac{1}{2} \\ I_{\frac{3}{2}} &\leq M\tau \left[e^{\tau^2 y} \tau + 2e^{\frac{y}{2}} \left(\frac{1}{\sqrt{2}} - \tau \right) + \frac{\sqrt{2}}{y} (e^y - e^{\frac{y}{2}}) \right], \text{ for } \tau < \frac{1}{\sqrt{2}} \\ I_{\frac{1}{2}} &\geq L(1)\tau \left[e^{\tau^2 y} \left(\frac{1}{\tau} - \frac{1}{\sqrt{\tau}} \right) + \frac{\sqrt{2}}{y} (e^y - e^{\tau y}) + \frac{1}{2y} (e^y - e^{\frac{y}{2}}) \right], \text{ for } \tau < \frac{1}{2}. \end{aligned}$$

Proof. Note that since L is nondecreasing over $(0, \infty)$ and $0 < \tau^2 < 1$, for each fixed $\tau \in (0, 1)$ we have, $L(t) \geq L(1)$ for all $t \geq \frac{1}{1-\tau^2} > 1$. Therefore,

$$\begin{aligned} I_{\frac{5}{2}} &= \int_0^\infty \frac{(t\tau^2)^2}{(1+t\tau^2)^{5/2}} t^{-3/2} L(t) e^{\frac{t\tau^2}{1+t\tau^2}y} dt \\ &= \frac{1}{\tau} \int_0^\infty \left(\frac{t\tau^2}{1+t\tau^2} \right)^{5/2} t^{-2} L(t) e^{\frac{t\tau^2}{1+t\tau^2}y} dt \\ &\geq \frac{L(1)}{\tau} \int_{\frac{1}{1-\tau^2}}^\infty \left(\frac{t\tau^2}{1+t\tau^2} \right)^{5/2} t^{-2} e^{\frac{t\tau^2}{1+t\tau^2}y} dt \end{aligned} \tag{A.36}$$

Now putting $u = \frac{t\tau^2}{1+t\tau^2}$ in (A.36) we obtain,

$$\begin{aligned} I_{\frac{5}{2}} &\geq L(1)\tau \int_{\tau^2}^1 u^{1/2} e^{uy} du \\ &= L(1)\tau \left[\frac{\tau}{y} (e^{y/2} - e^{\tau^2 y}) + \frac{1}{\sqrt{2}y} (e^y - e^{y/2}) \right], \text{ for } \tau < \frac{1}{\sqrt{2}} \end{aligned}$$

where the last equality follows using the proof of Lemma A.1 of van der Pas et al. (2014).

Next observe that

$$\begin{aligned} I_{\frac{1}{2}} &= \int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-3/2} L(t) e^{\frac{t\tau^2}{1+t\tau^2}y} dt \\ &= \frac{1}{\tau} \int_0^\infty \left(\frac{t\tau^2}{1+t\tau^2} \right)^{1/2} t^{-2} L(t) e^{\frac{t\tau^2}{1+t\tau^2}y} dt \\ &= \tau \int_0^1 u^{-3/2} L\left(\frac{1}{\tau^2} \frac{u}{1-u}\right) e^{uy} du \text{ [putting } u = \frac{t\tau^2}{1+t\tau^2}] \\ &= \tau \left[\int_0^{\tau^2} u^{-3/2} L\left(\frac{1}{\tau^2} \frac{u}{1-u}\right) e^{uy} du + \int_{\tau^2}^1 u^{-3/2} L\left(\frac{1}{\tau^2} \frac{u}{1-u}\right) e^{uy} du \right] \end{aligned} \tag{A.37}$$

Now observe that $e^{uy} \leq e^{\tau^2 y}$ for all $u \leq \tau^2$. Using this fact and applying the change of variable $t = \frac{1}{\tau^2} \frac{u}{1-u}$ in the first integral on the right hand side of (A.37) we obtain,

$$\begin{aligned}
 \int_0^{\tau^2} u^{-3/2} L\left(\frac{1}{\tau^2} \frac{u}{1-u}\right) e^{uy} du &\leq e^{\tau^2 y} \int_0^{\tau^2} u^{-3/2} L\left(\frac{1}{\tau^2} \frac{u}{1-u}\right) du \\
 &= \frac{e^{\tau^2 y}}{\tau} \int_0^{\frac{1}{1-\tau^2}} \frac{t^{-3/2}}{\sqrt{1+t\tau^2}} L(t) dt \\
 &\leq \frac{e^{\tau^2 y}}{\tau} \int_0^\infty t^{-3/2} L(t) dt \quad [\text{since } \frac{1}{\sqrt{1+t\tau^2}} \leq 1] \\
 &= \frac{K^{-1} e^{\tau^2 y}}{\tau}
 \end{aligned} \tag{A.38}$$

For the second integral on the right hand side of (A.37), we observe that the function L is bounded by the constant $M > 0$. Using this observation and then apply the same arguments given in the proof of Lemma A.1 of van der Pas et al. (2014), we obtain,

$$\int_{\tau^2}^1 u^{-3/2} L\left(\frac{1}{\tau^2} \frac{u}{1-u}\right) e^{uy} du \leq 2M \left[e^{\tau y} \left(\frac{1}{\tau} - \frac{1}{\sqrt{\tau}} \right) + e^{\frac{y}{2}} \left(\frac{1}{\sqrt{\tau}} - \sqrt{2} \right) + \frac{\sqrt{2}}{y} (e^y - e^{\frac{y}{2}}) \right] \text{ for } \tau < \frac{1}{2}. \tag{A.39}$$

(A.37), (A.38) and (A.39) together immediately give the stated upper bound on $I_{\frac{1}{2}}$.

Again note that since $\tau^2 < u < 1$ we have $\frac{1}{\tau^2} \frac{u}{1-u} > \frac{1}{1-\tau^2} > 1$. Hence $L\left(\frac{1}{\tau^2} \frac{u}{1-u}\right) \geq L\left(\frac{1}{1-\tau^2}\right) \geq L(1)$ as L is nondecreasing. Using this observation and (A.37) and then applying the same reasoning given in the proof of Lemma A.1 of van der Pas et al. (2014), we obtain,

$$\begin{aligned}
 I_{\frac{1}{2}} &\geq \tau \int_{\tau^2}^1 u^{-3/2} L\left(\frac{1}{\tau^2} \frac{u}{1-u}\right) e^{uy} du \\
 &\geq L(1) \tau \int_{\tau^2}^1 u^{-3/2} e^{uy} du \\
 &= L(1) \tau \left[e^{\tau^2 y} \left(\frac{1}{\tau} - \frac{1}{\sqrt{\tau}} \right) + \frac{\sqrt{2}}{y} (e^y - e^{\tau y}) + \frac{1}{2y} (e^y - e^{\frac{y}{2}}) \right], \text{ for } \tau < \frac{1}{2}.
 \end{aligned}$$

The stated upper bound for $I_{\frac{3}{2}}$ follows immediately by noting that L is bounded by the constant $M > 0$ and subsequently by change the variable $u = t\tau^2/(1+t\tau^2)$ followed by the arguments used in the proof of Lemma A.1 of van der Pas et al. (2014), thereby completing the proof of Lemma A.6. \blacksquare

Proof of Theorem 3.4

Proof. From (A.24) we have,

$$\begin{aligned}
 \text{Var}(\theta|x) &\geq x^2 E \left[(1 - \kappa)^2 |x, \tau \right] - x^2 E^2 \left[(1 - \kappa) |x, \tau \right] \\
 &= x^2 \left[\frac{\int_0^\infty \frac{(t\tau^2)^2}{(1+t\tau^2)^{5/2}} t^{-3/2} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-3/2} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt} - \left(\frac{\int_0^\infty \frac{t\tau^2}{(1+t\tau^2)^{3/2}} t^{-3/2} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_0^\infty \frac{1}{(1+t\tau^2)^{1/2}} t^{-3/2} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt} \right)^2 \right] \\
 &= 2y \left[\frac{I_{\frac{5}{2}}}{I_{\frac{1}{2}}} - \left(\frac{I_{\frac{3}{2}}}{I_{\frac{1}{2}}} \right)^2 \right]
 \end{aligned}$$

by putting $y = \frac{x^2}{2}$ and rest of the proof follows by applying Lemma A.6 and the same set of arguments given in the proof of Theorem 3.4 of van der Pas et al. (2014). ■

Lemma A.7. *Let us fix any $0 < \eta < 1$ and any $0 < \delta < 1$. Consider the general class of shrinkage priors where the prior distribution for the local shrinkage parameters $\pi(\lambda_i^2)$ is given by (2.5) with $a \in (0, 1)$ and the corresponding slowly varying component $L(\cdot)$ satisfies Assumption 2.1. Then for each fixed $x \in \mathbb{R}$ and every fixed $0 < \tau < 1$, the corresponding posterior shrinkage coefficients $E(\kappa|x, \tau)$ can be bounded above by a real valued function $g(x, \tau)$, depending on η and δ , and is given by,*

$$g(x, \tau) = \begin{cases} C_* \left[\left| x^2 \int_0^{sx^2} e^{-u/2} u^{a+1/2-1} du \right| \right]^{-1} + \frac{H(a, \eta, \delta) e^{-\frac{\eta(1-\delta)x^2}{2}}}{\tau^{2a} \Delta(\tau^2, \eta, \delta)}, & \text{if } |x| > 0, \\ 1, & \text{if } x = 0, \end{cases} \quad (\text{A.40})$$

for some global constant $C_* \equiv C_*(a, \eta, L) > 0$ which is independent of both x and τ . The function $g(x, \tau)$ defined in (A.40) above satisfies the following:

Given any $\zeta > \frac{2}{\eta(1-\delta)}$,

$$\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\zeta \log(\frac{1}{\tau^{2a}})}} g(x, \tau) = 0.$$

Proof. Proof follows using the same set of arguments as in the proof of Lemma A.3. ■

Remark A.1. For each fixed τ , the function $g(x, \tau)$ defined in (A.40) has exactly one point of discontinuity at the origin and hence it is measurable.

Proof of Theorem 4.1

Proof. First of all, it should be observed that, for each i , $X_i \sim N(0, 1)$ under H_{0i} , and hence the corresponding type I error probability does not depend on i . Let t_1 denote the common value of t_{1i} 's, $i = 1, \dots, n$. Hence, by definition,

$$t_1 = \Pr(E(1 - \kappa_1 | X_1, \tau) > 0.5 | H_{01} \text{ is true}) \quad (\text{A.41})$$

Now, the upper bound for t_1 as given in the statement of Theorem 4.1, is a simple consequence of Theorem 6 of Ghosh et al. (2015) with some appropriately chosen finite positive constant $H(a, \eta, \delta)$

which is independent of both i and n . Hence we omit the proof. However, the corresponding proof for the lower bound of t_1 require some novel arguments as given below.

Let us fix any $\eta \in (0, 1)$ and any $\delta \in (0, 1)$. Then, by Lemma A.7, for each fixed $\tau < 1$, the function $E(\kappa|x, \tau)$ is bounded above by the function $g(x, \tau)$, where the function $g(\cdot, \tau)$ has already been defined in (A.40). Hence, for each fixed $\tau < 1$ and for every ω in the sample space, we have the following:

$$\begin{aligned} \left\{ \omega : E(1 - \kappa_1|X_1(\omega), \tau) > 0.5 \right\} &= \left\{ \omega : E(\kappa_1|X_1(\omega), \tau) < 0.5 \right\} \\ &\supseteq \left\{ \omega : g(X_1(\omega), \tau) < 0.5 \right\} \equiv B_n^c, \text{ say,} \end{aligned} \quad (\text{A.42})$$

where $B_n \equiv \{\omega : g(X_1(\omega), \tau) \geq 0.5\}$.

Let us fix any $\zeta > \frac{2}{\eta(1-\delta)}$ and define the event C_n as $C_n \equiv \left\{ \omega : |X_1(\omega)| > \sqrt{\zeta \log(\frac{1}{\tau^{2a}})} \right\}$.

Then using (A.46) and (A.42), we have,

$$\begin{aligned} t_1 &= \Pr(E(\kappa_1|X_1, \tau) < 0.5 | H_{01} \text{ is true}) \\ &\geq \Pr(g(X_1, \tau) < 0.5 | H_{01} \text{ is true}) \\ &= \Pr(B_n^c | H_{01} \text{ is true}) \\ &\geq \Pr(B_n^c \cap C_n | H_{01} \text{ is true}) \\ &= \Pr(B_n^c | C_n, H_{01} \text{ is true}) \Pr(C_n | H_{01} \text{ is true}) \\ &= [1 - \Pr(B_n | C_n, H_{01} \text{ is true})] \Pr(C_n | H_{01} \text{ is true}). \end{aligned} \quad (\text{A.43})$$

Recall that the function $g(x, \tau)$ is measurable, non-negative and is continuously decreasing in $|x|$ for $|x| \neq 0$. Hence $E(g(X_1, \tau) | |X_1| > \sqrt{\zeta \log(1/\tau^{2a})}, H_{01} \text{ is true})$ is well defined and is bounded for all sufficiently small $\tau \in (0, 1)$. Using these facts and applying Markov's inequality, we obtain for all sufficiently $\tau < 1$, the following:

$$\begin{aligned} \Pr(B_n | C_n, H_{01} \text{ is true}) &= \Pr(g(X_1, \tau) \geq 0.5 | |X_1| > \sqrt{\zeta \log(1/\tau^{2a})}, H_{01} \text{ is true}) \\ &\leq 2E(g(X_1, \tau) | |X_1| > \sqrt{\zeta \log(1/\tau^{2a})}, H_{01} \text{ is true}) \\ &\leq 2 \sup_{|x| > \sqrt{\zeta \log(\frac{1}{\tau^{2a}})}} g(x, \tau) \\ &\rightarrow 0 \text{ as } \tau \rightarrow 0. \end{aligned}$$

Since $\tau \rightarrow 0$ as $n \rightarrow \infty$, we have,

$$\lim_{n \rightarrow \infty} \Pr(B_n | C_n, H_{01} \text{ is true}) = 0,$$

whence it follows

$$\Pr(B_n^c | C_n, H_{01} \text{ is true}) = 1 - o(1) \text{ as } n \rightarrow \infty. \quad (\text{A.44})$$

Again, noting that under H_{01} , $X_1 \stackrel{d}{=} Z$, we have for all sufficiently small $\tau < 1$, the following:

$$\begin{aligned}
\Pr(C_n | H_{01} \text{ is true}) &= \Pr(|Z| > \sqrt{\zeta \log(1/\tau^{2a})}) \\
&= 2 \Pr(Z > \sqrt{\zeta \log(1/\tau^{2a})}) \\
&\geq 2 \frac{\phi(\sqrt{\zeta \log(1/\tau^{2a})})}{\sqrt{\zeta \log(1/\tau^{2a})}} \left(1 - \frac{1}{\zeta \log(1/\tau^{2a})}\right) [\text{using Mill's ratio}] \\
&\geq 2 \frac{\phi(\sqrt{\zeta \log(1/\tau^{2a})})}{\sqrt{\zeta \log(1/\tau^{2a})}} \left(1 - \frac{1}{2 \log(1/\tau^{2a})}\right) [\text{since } \zeta > 2] \\
&= G(a, \eta, \delta) \frac{(\tau^{2a})^{\frac{\zeta}{2}}}{\sqrt{\log(\frac{1}{\tau^2})}} (1 + o(1)) \text{ as } n \rightarrow \infty.
\end{aligned} \tag{A.45}$$

On combining (A.43), (A.44) and (A.45), the stated result follows immediately. \blacksquare

Proof of Theorem 4.2

Proof. First observe that, for each i , $X_i \sim N(0, 1 + \psi^2)$ under H_{Ai} , and hence the corresponding type II error probability does not depend on i . Let t_1 denote the common value of t_{2i} 's, $i = 1, \dots, n$. Hence, by definition,

$$t_2 = \Pr(E(1 - \kappa_1 | X_1, \tau) \leq 0.5 | H_{A1} \text{ is true}) \tag{A.46}$$

Now, the lower bound for t_2 as given in the statement of Theorem 4.2, is a simple consequence of Theorem 8 of Ghosh et al. (2015). Hence we skip the proof of this part. However, the corresponding proof for the upper bound of t_2 require some novel arguments as given below.

Let us fix any $\eta \in (0, 1)$ and any $\delta \in (0, 1)$. Choose any $\zeta > \frac{2}{\eta(1-\delta)}$. Then using Lemma A.7, it follows that, for each fixed $\tau < 1$, the function $E(\kappa | x, \tau)$ is bounded above by the function $g(x, \tau)$, where the function $g(\cdot, \tau)$ has already been defined in (A.40.) Hence, for each fixed $\tau < 1$, we have the following:

$$\begin{aligned}
t_2 &= \Pr(E(\kappa_1 | X_1, \tau) \geq 0.5 | H_{A1} \text{ is true}) \\
&\leq \Pr(g(X_1, \tau) \geq 0.5 | H_{A1} \text{ is true}) \\
&= \Pr(B_n | H_{A1} \text{ is true}) \\
&= \Pr(B_n \cap C_n | H_{A1} \text{ is true}) + \Pr(B_n \cap C_n^c | H_{A1} \text{ is true}) \\
&= \Pr(B_n | C_n, H_{A1} \text{ is true}) \Pr(C_n | H_{A1} \text{ is true}) + \Pr(B_n \cap C_n^c | H_{A1} \text{ is true}) \\
&\leq \Pr(B_n | C_n, H_{A1} \text{ is true}) + \Pr(C_n^c | H_{A1} \text{ is true}).
\end{aligned} \tag{A.47}$$

where the events B_n and C_n are already defined in the proof of Theorem 4.1.

Now applying the same reasoning as in the proof of Theorem 4.1 it follows that

$$\lim_{n \rightarrow \infty} \Pr(B_n | C_n, H_{A1} \text{ is true}) = 0. \tag{A.48}$$

Again, since $0 < \lim_{n \rightarrow \infty} \frac{\tau}{p^\alpha} < \infty$, for $\alpha > 0$, using condition (iii) of Assumption 4.1, it follows that $\lim_{n \rightarrow \infty} \log(\frac{1}{\tau^2})/\psi^2 = \alpha C$. Thus, for all sufficiently large n , we have the following:

$$\begin{aligned}
 \Pr(C_n^c | H_{A1} \text{ is true}) &= \Pr\left(|X_1| \leq \sqrt{\zeta \log(\frac{1}{\tau^{2a}})} | H_{A1} \text{ is true}\right) \\
 &= \Pr\left(|Z| \leq \sqrt{\zeta a} \sqrt{\frac{\log(\frac{1}{\tau^2})}{1 + \psi^2}}\right) \\
 &= \Pr\left(|Z| \leq \sqrt{\zeta a} \sqrt{\frac{\log(\frac{1}{\tau^2})}{\psi^2}} (1 + o(1))\right) \text{ as } n \rightarrow \infty \\
 &= \Pr(|Z| \leq \sqrt{\zeta a \alpha \sqrt{C}} (1 + o(1))) \text{ as } n \rightarrow \infty \\
 &= [2\Phi(\sqrt{\zeta a \alpha \sqrt{C}}) - 1](1 + o(1)) \text{ as } n \rightarrow \infty.
 \end{aligned} \tag{A.49}$$

Combining (A.47), (A.48) and (A.49), it therefore follows that

$$t_2 \leq [2\Phi(\sqrt{\zeta a \alpha \sqrt{C}}) - 1](1 + o(1)),$$

for all sufficiently large n , thereby completing the proof of Theorem 4.2. ■

Proof of Theorem 4.5

Proof. Let us first fix any $\eta \in (0, 1)$ and any $\delta \in (0, 1)$. Then combining the results of Theorem 4.1 and Theorem 4.2, together with (4.8) and the arguments employed for proving Theorem 1 of Ghosh et al. (2015), it follows easily that, that the Bayes risk R_{OG} of the induced decisions, (4.10) based on the general class of one-group priors under study, with $a \in [0.5, 1)$, satisfies the following:

$$\frac{2\Phi(\sqrt{2a\alpha\sqrt{C}}) - 1}{2\Phi(\sqrt{C}) - 1} (1 + o(1)) \leq \frac{R_{OG}}{R_{Opt}^{BO}} \leq \frac{2\Phi(\sqrt{\zeta a \alpha \sqrt{C}}) - 1}{2\Phi(\sqrt{C}) - 1} (1 + o(1)) \text{ as } n \rightarrow \infty, \tag{A.50}$$

for any arbitrary but fixed $\zeta > \frac{2}{\eta(1-\delta)}$, where the $o(1)$ terms depend on choices of ζ , η and δ only. Then taking limit inferior and limit superior in (A.50) as $n \rightarrow \infty$, it follows

$$\frac{2\Phi(\sqrt{2a\alpha\sqrt{C}}) - 1}{2\Phi(\sqrt{C}) - 1} \leq \liminf_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} \leq \limsup_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} \leq \frac{2\Phi(\sqrt{\zeta a \alpha \sqrt{C}}) - 1}{2\Phi(\sqrt{C}) - 1} \tag{A.51}$$

for each fixed $\eta \in (0, 1)$ and each fixed $\delta \in (0, 1)$ and for any $\zeta > 2/(\eta(1 - \delta))$.

Now observe that the multiple testing rules under study do not depend on how $\eta \in (0, 1)$, $\delta \in (0, 1)$ and $\zeta > 2/(\eta(1 - \delta))$ are chosen. Hence the ratio R_{OG}/R_{Opt}^{BO} is free of any $\eta, \delta \in (0, 1)$ and any $\zeta > 2/(\eta(1 - \delta))$, for all $n \geq 1$. Thus, the limit inferior and the limit superior terms in (A.51) are also independent of the choices of η, δ and ζ . But $\zeta > 2/(\eta(1 - \delta))$ in (A.51) is arbitrary. Therefore, taking infimum over all such ζ 's in (A.51) and using the continuity of $\Phi(\cdot)$, we obtain,

$$\frac{2\Phi(\sqrt{2a\alpha\sqrt{C}}) - 1}{2\Phi(\sqrt{C}) - 1} \leq \liminf_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} \leq \limsup_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} \leq \frac{2\Phi(\sqrt{\frac{2a\alpha}{\eta(1-\delta)}}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1} \tag{A.52}$$

Once again (A.52) holds for every fixed $\eta \in (0, 1)$ and every fixed $\delta \in (0, 1)$. Hence, using the preceding discussion and taking infimum in (A.52) over all possible choices of $(\eta, \delta) \in (0, 1) \times (0, 1)$ and using the continuity of $\Phi(\cdot)$, we finally obtain

$$\frac{2\Phi(\sqrt{2a\alpha}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1} \leq \liminf_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} \leq \limsup_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} \leq \frac{2\Phi(\sqrt{2a\alpha}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1}$$

whence we have

$$\lim_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} = \frac{2\Phi(\sqrt{2a\alpha}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1}.$$

That $\lim_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} = 1$ for $a = 0.5$ and $\alpha = 1$ is obvious. This completes the proof of Theorem 4.5. ■

Acknowledgements

The authors would like to thank Professor Jayanta Kumar Ghosh for letting them aware about the work of van der Pas et al. (2014) on the posterior contraction properties of the horseshoe prior.

References

- Armagan, A., Dunson, D. B., and Clyde, M. (2011). Generalized beta mixtures of gaussians. In Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. C. N., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 24, pages 523–531.
- Armagan, A., Dunson, D. B., and Lee, J. (2012). Generalized double pareto shrinkage. *Statistica Sinica*, 23(1):119–143.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300.
- Bhattacharya, A., Pati, D., Pillai, N., and Dunson, D. B. (2012). Bayesian shrinkage. *arXiv:1212.6088v1*.
- Bhattacharya, A., Pati, D., Pillai, N., and Dunson, D. B. (2014). Dirichlet-laplace priors for optimal shrinkage. *arXiv:1401.5398v1*.
- Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37:1705–1732.
- Bogdan, M., Chakrabarti, A., Frommlet, F., and Ghosh, J. K. (2011). Asymptotic bayes-optimality under sparsity of some multiple testing procedures. *The Annals of Statistics*, 39(3):1551–1579.
- Bogdan, M., Ghosh, J. K., and Tokdar, S. T. (2008). A comparison of the benjamini-hochberg procedure with some bayesian rules for multiple testing. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, volume 1, pages 211–230, Beachwood, OH. IMS Collections, IMS.
- Carvalho, C., Polson, N., and Scott, J. (2009). Handling sparsity via the horseshoe. *Journal of Machine Learning Research W&CP*, 5:73–80.
- Carvalho, C., Polson, N., and Scott, J. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.
- Castillo, I., Schmidt-Heiber, J., and van der Vaart, A. W. (2014). Bayesian linear regression with sparse priors. *arXiv:1403.0735*.

- Castillo, I. and van der Vaart, A. W. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *The Annals of Statistics*, 40(4):2069–2101.
- Datta, J. and Ghosh, J. K. (2013). Asymptotic properties of bayes risk for the horseshoe prior. *Bayesian Analysis*, 8(1):111–132.
- Donoho, D. L., Johnstone, I. M., Hoch, J. C., and Stern, A. S. (1992). Maximum entropy and the nearly black object (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 54:41–81.
- Efron, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association*, 99(465):96–104.
- Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). Convergence rates of posterior distribution. *The Annals of Statistics*, 28(2):500–531.
- Ghosh, P., Tang, X., Ghosh, M., and Chakrabarti, A. (2015). Asymptotic properties of bayes risk of a general class of shrinkage priors in multiple hypothesis testing under sparsity. *Bayesian Analysis (Advance Publications)*.
- Griffin, J. E. and Brown, P. J. (2005). Alternative prior distributions for variable selection with very many more variables than observations. Technical report, University of Warwick.
- Griffin, J. E. and Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188.
- Griffin, J. E. and Brown, P. J. (2012). Structuring shrinkage: some correlated priors for regression. *Biometrika*, 99(2):481–487.
- Griffin, J. E. and Brown, P. J. (2013). Some priors for sparse regression modeling. *Bayesian Analysis*, 8(3):691–702.
- Hans, C. (2009). Bayesian lasso regression. *Biometrika*, 96(4):835–845.
- Jiang, W. and Zhang, C. H. (2009). General maximum likelihood empirical bayes estimation of normal means. *The Annals of Statistics*, 37(2):1647–1684.
- Johnstone, I. and Silverman, B. W. (2004). Needles and straw in haystacks: Empirical-bayes estimates of possibly sparse sequences. *The Annals of Statistics*, 32(4):1594–1649.
- Martin, R. and Walker, S. G. (2014). Asymptotically minimax empirical bayes estimation of a sparse normal mean vector. *Electronic Journal of Statistics*, 8:2188–2206.
- Mitchell, T. and Beauchamp, J. (1988). Bayesian variable selection in linear regression (with discussion). *Journal of the American Statistical Association*, 83(404):1023–1036.
- Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.
- Polson, N. G. and Scott, J. G. (2011). Shrink globally, act locally: Sparse bayesian regularization and prediction. In *Bayesian Statistics 9, Proceedings of the 9th Valencia International Meeting*, pages 501–538. Oxford University Press.
- Polson, N. G. and Scott, J. G. (2012). On the half-cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(2):1–16.
- Tipping, M. (2001). Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244.
- van der Pas, S. L., Kleijn, B. J. K., and van der Vaart, A. W. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics*, 8:2585–2618.
- Yuan, M. and Lin, Y. (2005). Efficient empirical bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, 100:1215–1225.